

09/890, 806.

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compen Ltd.

OM protein - protein search, using sw model

Run on: January 22, 2004, 11:37:25 ; Search time 21.7747 Seconds  
(without alignments)  
225.401 Million cell updates/sec

Title: US-09-890-806-3\_COPY\_28\_143  
Perfect score: 629  
Sequence: 1 AGEDALRPWKSTAKHPWFQI.....LQGVIMLRCPBELRDVYTS 116

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA:  
1: /cgn2\_6/prodata/2/1aa/5A\_COMB.pep:\*  
2: /cgn2\_6/prodata/2/1aa/5B\_COMB.pep:\*  
3: /cgn2\_6/prodata/2/1aa/6A\_COMB.pep:\*  
4: /cgn2\_6/prodata/2/1aa/6B\_COMB.pep:\*  
5: /cgn2\_6/prodata/2/1aa/6CTOS\_COMB.pep:\*  
6: /cgn2\_6/prodata/2/1aa/Backfilltest.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	77.5	12.3	676	3	US-08-947-965-71 Sequence 71, Appl
2	73.5	11.7	729	4	US-09-252-991A-32535 Sequence 32535, A
3	70.5	11.2	675	3	US-08-947-965-76 Sequence 76, Appl
4	69	11.0	893	4	US-09-514-302-4 Sequence 4, Appl
5	69	11.0	1338	4	US-09-514-302-2 Sequence 2, Appl
6	68.5	10.9	640	4	US-09-252-991A-27542 Sequence 27542, A
7	67	10.7	240	4	US-09-152-060-100 Sequence 100, App
8	67	10.7	335	4	US-09-152-060-64 Sequence 64, Appl
9	67	10.7	391	3	US-08-706-216-6 Sequence 6, Appl
10	66.5	10.6	100	4	US-09-430-702-1 Sequence 1, Appl
11	66	10.5	790	3	US-08-960-780-4 Sequence 4, Appl
12	66	10.5	790	3	US-09-073-898-4 Sequence 4, Appl
13	65.5	10.4	516	4	US-08-868-373-14 Sequence 14, Appl
14	65	10.3	587	4	US-09-328-352-8096 Sequence 8096, Ap
15	64.5	10.3	310	4	US-09-252-991A-17381 Sequence 17381, A
16	64	10.2	746	2	US-08-838-2198-6 Sequence 6, Appl
17	64	10.2	746	3	US-09-233-336A-6 Sequence 6, Appl
18	64	10.2	746	3	US-09-233-752A-6 Sequence 6, Appl
19	64	10.2	746	3	US-09-402-036-6 Sequence 6, Appl
20	64	10.2	746	3	US-09-402-036-6 Sequence 6, Appl
21	64	10.2	789	1	US-08-471-033-32 Sequence 29, Appl
22	64	10.2	789	1	US-08-471-033-32 Sequence 29, Appl
23	64	10.2	789	2	US-08-471-044-29 Sequence 29, Appl
24	64	10.2	789	2	US-08-471-044-32 Sequence 32, Appl
25	64	10.2	789	2	US-08-463-483A-29 Sequence 32, Appl
26	64	10.2	789	2	US-08-463-483A-32 Sequence 32, Appl
27	64	10.2	789	2	US-08-471-046A-29 Sequence 29, Appl

28	64	10.2	789	2	US-08-471-046A-32 Sequence 32, Appl
29	64	10.2	789	2	US-08-470-566B-29 Sequence 29, Appl
30	64	10.2	789	2	US-08-470-566B-32 Sequence 32, Appl
31	64	10.2	789	2	US-08-838-219B-2 Sequence 2, Appl
32	64	10.2	789	2	US-08-838-219B-4 Sequence 4, Appl
33	64	10.2	789	2	US-08-469-334-29 Sequence 29, Appl
34	64	10.2	789	2	US-08-469-334-32 Sequence 32, Appl
35	64	10.2	789	2	US-09-300-529-29 Sequence 29, Appl
36	64	10.2	789	3	US-09-300-529-32 Sequence 32, Appl
37	64	10.2	789	3	US-09-233-336A-2 Sequence 2, Appl
38	64	10.2	789	3	US-09-233-336A-4 Sequence 4, Appl
39	64	10.2	789	3	US-09-233-752A-2 Sequence 2, Appl
40	64	10.2	789	3	US-09-233-752A-4 Sequence 4, Appl
41	64	10.2	789	3	US-09-402-036-2 Sequence 2, Appl
42	64	10.2	789	3	US-09-402-036-4 Sequence 4, Appl
43	64	10.2	789	4	US-09-002-285-78 Sequence 78, Appl
44	64	10.2	789	4	US-09-002-285-80 Sequence 80, Appl
45	64	10.2	789	4	US-09-002-285-92 Sequence 92, Appl

ALIGNMENTS

RESULT 1  
US-08-947-965-71  
; Sequence 71, Application US/08947965A  
; Patent No. 6004790  
; GENERAL INFORMATION:  
; APPLICANT: Dijkhuizen, Lubbert  
; APPLICANT: Dijkhuizen, Bauke  
; APPLICANT: Andersen, Carsten  
; APPLICANT: Oosten, Claus von der  
; TITLE OF INVENTION: Cyclomaltoextrin Glucanotransferase  
; FILE REFERENCE: 4285.204-US  
; CURRENT APPLICATION NUMBER: US/08/947,965A  
; EARLIER FILING DATE: 1997-10-09  
; EARLIER APPLICATION NUMBER: 0477/95  
; EARLIER FILING DATE: 1995-04-21  
; EARLIER APPLICATION NUMBER: 1173/95  
; EARLIER FILING DATE: 1995-10-17  
; EARLIER APPLICATION NUMBER: 1281/95  
; EARLIER FILING DATE: 1995-11-16  
; EARLIER APPLICATION NUMBER: PCT/DK96/00179  
; EARLIER FILING DATE: 1996-04-22  
; NUMBER OF SEQ ID NOS: 78  
; SOFTWARE: FaSTSeq for Windows Version 3.0  
; SEQ ID NO 71  
; LENGTH: 676  
; TYPE: PRT  
; ORGANISM: Bacillus sp.  
US-08-947-965-71  
Query Match 12.3%; Score 77.5; DB 3; Length 676;  
Best local similarity 24.0%; Pred. No. 0.67;  
Matches 25; Conservative 22; Mismatches 44; Indels 13; Gaps 3;  
QY 12 TAKHPFQIEDKRCYINDGKLFARGSIYGNMGRFVDFPADYGVGGENLYVADYEFVP 71  
DB 132 TNNHSPALETPNVYENGAIDNGALLGN-----YSNDQNLFHHNGTDFSS 180  
QY 72 GE-SLKMYVRNLDWPIFETLALRLVLOGDVIMLRCPBELRDV 113  
DB 181 YEDSYRNLVDLADYDLNNTVMDYIKESIKFMDIGIGIRVD 224  
RESULT 2  
US-09-252-991A-32535  
; Sequence 32535, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

;; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
;; FILE REFERENCE: 107196.136  
;; CURRENT APPLICATION NUMBER: US/09/252.991A  
;; CURRENT FILING DATE: 1999-02-18  
;; PRIOR APPLICATION NUMBER: US 60/074,788  
;; PRIOR FILING DATE: 1998-02-18  
;; PRIOR APPLICATION NUMBER: US 60/094,190  
;; PRIOR FILING DATE: 1998-07-27  
;; NUMBER OF SEQ ID NOS: 33142  
;; SEQ ID NO 32535  
;; LENGTH: 729  
;; TYPE: PRT  
;; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-32535

Query Match 11.7%; Score 73.5; DB 4; Length 729;  
Best Local Similarity 29.3%; Pred. No. 2.3;  
Matches 34; Conservative 16; Mismatches 41; Indels 25; Gaps 7;

QY 1 AGED-----ALRPWKSRAKHPFOIE-DNRCYIDNGKLFARGSIYGNMSRFVDPKADYG 54  
DB 559 AGEDLGVALVDPKRTHTVRALAOVDADLAVGI-----RAGGVDEHRRVFPAAEGG-G 611  
QY 55 GVGENTLYHAD-DVEFVPGESLKNVNVNLDVMPFIFF---TLAURLVLOGDVIMLRG 106  
DB 612 GVGEGDEPHRDADYVARAAGQ-----VDLARVERLRRLAVELLKAGDVILLRC 659

RESULT 3  
US-08-947-965-76  
; Sequence 76, Application US/08947965A  
; Patent No. 6004790  
; GENERAL INFORMATION:  
; APPLICANT: Dijkhuizen, Lubbert  
; APPLICANT: Dijkstra, Bauke  
; APPLICANT: Andersen, Carsten  
; APPLICANT: Ostren, Claus von der  
; TITLE OF INVENTION: Cyclomaltohextrin Glucanotransferase  
; FILE OF INVENTION: Variants  
; FILE REFERENCE: 4285.204-US  
; CURRENT APPLICATION NUMBER: US/08/947,965A  
; CURRENT FILING DATE: 1997-10-09  
; EARLIER APPLICATION NUMBER: 0477/95  
; EARLIER FILING DATE: 1995-04-21  
; EARLIER APPLICATION NUMBER: 1173/95  
; EARLIER FILING DATE: 1995-10-17  
; EARLIER APPLICATION NUMBER: 1281/95  
; EARLIER FILING DATE: 1995-11-16  
; EARLIER APPLICATION NUMBER: PCT/DK96/00179  
; EARLIER FILING DATE: 1996-04-22  
; NUMBER OF SEQ ID NOS: 78  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 76  
; LENGTH: 675  
; TYPE: PRT  
; ORGANISM: Bacillus ohbensis  
US-08-947-965-76

Query Match 11.2%; Score 70.5; DB 3; Length 675;  
Best Local Similarity 25.0%; Pred. No. 4.8;  
Matches 26; Conservative 21; Mismatches 44; Indels 13; Gaps 4;

QY 12 TAKHPWFOIEDNRCYIDNGKLFARGSIYGNMSRFVDPKADYGGVGNLYVHADVFPV 71  
DB 130 TPNHSSPALLETDPYARNGAVYNDGVLIGNYSN---DP-----NHLFHHNGGTDFSS 178  
QY 72 GE-SLKNNVRLVDMPFIFFETLRLVLOGDVIMLRGPELRVD 113  
DB 179 YEDSIYRLVLDYDLANTVMDQYLKESITKJMDKIDGIRVD 222

RESULT 4  
US-09-514-302-4

;; Sequence 4, Application US/09514302  
;; Patent No. 6338959  
;; GENERAL INFORMATION:  
;; APPLICANT: HATADA, Yuji  
;; APPLICANT: IGARASHI, Kazuaki  
;; APPLICANT: OZAKI, Katsuya  
;; APPLICANT: ARA, Katsutoshi  
;; APPLICANT: KAWAI, Shuji  
;; APPLICANT: ITO, Susumu  
;; TITLE OF INVENTION: GENE FOR ENZYME HAVING BOTH ALKALINE PHOSPHATASE AND  
;; FILE REFERENCE: 2173-105P  
;; CURRENT APPLICATION NUMBER: US/09/514,302  
;; CURRENT FILING DATE: 2000-02-28  
;; EARLIER APPLICATION NUMBER: 08/952,084  
;; EARLIER FILING DATE: 1997-11-10  
;; NUMBER OF SEQ ID NOS: 14  
;; SOFTWARE: PatentIn Ver. 2.1  
;; SEQ ID NO 4  
;; LENGTH: 893  
;; TYPE: PRT  
;; ORGANISM: Bacillus sp.  
US-09-514-302-4

Query Match 11.0%; Score 69; DB 4; Length 893;  
Best Local Similarity 32.0%; Pred. No. 11;  
Matches 24; Conservative 11; Mismatches 30; Indels 10; Gaps 3;

QY 12 TAKHPWFOIEDNRCYIDNGKLFARGSIYGNMSRFVDPKAD-----YGGVGNLYVHAD 65  
DB 82 SVVSGWKLIDEMVAY--DGLGALHEDGTATLKWSPKADNVSVVLVDKVDQNEVV--D 137  
QY 66 DVEFVPGESLKNVNR 80  
DB 138 TIEMVKGDRGWSVK 152

RESULT 5  
US-09-514-302-2  
; Sequence 2, Application US/09514302  
; Patent No. 6338959  
; GENERAL INFORMATION:  
; APPLICANT: HATADA, Yuji  
; APPLICANT: IGARASHI, Kazuaki  
; APPLICANT: OZAKI, Katsuya  
; APPLICANT: ARA, Katsutoshi  
; APPLICANT: KAWAI, Shuji  
; APPLICANT: ITO, Susumu  
; TITLE OF INVENTION: GENE FOR ENZYME HAVING BOTH ALKALINE PHOSPHATASE AND  
; FILE REFERENCE: 2173-105P  
; CURRENT APPLICATION NUMBER: US/09/514,302  
; CURRENT FILING DATE: 2000-02-28  
; EARLIER APPLICATION NUMBER: 08/952,084  
; EARLIER FILING DATE: 1997-11-10  
; NUMBER OF SEQ ID NOS: 14  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 1938  
; TYPE: PRT  
; ORGANISM: Bacillus sp.  
US-09-514-302-2

Query Match 11.0%; Score 69; DB 4; Length 1938;  
Best Local Similarity 32.0%; Pred. No. 31;  
Matches 24; Conservative 11; Mismatches 30; Indels 10; Gaps 3;

QY 12 TAKHPWFOIEDNRCYIDNGKLFARGSIYGNMSRFVDPKAD-----YGGVGNLYVHAD 65  
DB 1127 SVVSGWKLIDEMVAY--DGLGALHEDGTATLKWSPKADNVSVVLVDKVDQNEVV--D 1182  
QY 66 DVEFVPGESLKNVNR 80



RESULT 9  
US-08-706-216-6  
; Sequence 6, Application US/08706216  
; Patent No. 6140098  
; GENERAL INFORMATION:  
; APPLICANT: Balasubramanian, Srilam  
; APPLICANT: Ford, John  
; APPLICANT: Gorman, Daniel M.  
; APPLICANT: Zurawski, Gerard  
; TITLE OF INVENTION: MAMMALIAN PROTEASES; RELATED REAGENTS  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: DNA Research Institute  
; STREET: 901 California Avenue  
; CITY: Palo Alto  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94304-1104  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/706,216  
; FILING DATE: 30-AUG-1996  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ching, Edwin P.  
; REGISTRATION NUMBER: 34,090  
; REFERENCE/DOCKET NUMBER: DX0613  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-952-9196  
; TELEFAX: 415-496-1200  
; INFORMATION FOR SEQ ID NO: 6:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 391 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-706-216-6  
Query Match 10.7%; Score 67; DB 3; Length 391;  
Best Local Similarity 26.4%; Pred. No. 6.2;  
Matches 29; Conservative 18; Mismatches 35; Indels 28; Gaps 6;  
QY 14 KHPWFQIED-----NRGVINDGKLFARGSIYGNMRFVDPKADYGGVGEN-----59  
DB 151 RHAVYQAEHLIQTGTGCVSDSL---GSLIGPRTAVFRPR-----GDSLPRETRRV 202  
QY 60 -LYVHADVER-VPGESIKMNVNLDVW---PIFETLALRLVYQGVIM 103  
DB 203 ELYVVVNABEQMLGSEAAVHRVLEVVNHNVDKLYQKLNFRVLVVGELI 252  
RESULT 10  
US-09-430-702-1  
; Sequence 1, Application US/09430702  
; Patent No. 6472192  
; GENERAL INFORMATION:  
; APPLICANT: SCHULZ, GEORG E.  
; APPLICANT: PARISEGLA, GOETZ  
; APPLICANT: CANDUSSIO, ANTON  
; APPLICANT: WICH, GUNTER  
; TITLE OF INVENTION: CYCLODEXTRIN GLYCOSYL TRANSFERASES FOR PRODUCING  
; TITLE OF INVENTION: Y-CYCLODEXTRIN  
; FILE REFERENCE: SCHULZ-W2 CIP  
; CURRENT APPLICATION NUMBER: US/09/430,702  
; CURRENT FILING DATE: 1999-10-29  
; EARLIER APPLICATION NUMBER: 08/816,317  
; EARLIER FILING DATE: 1997-03-13

; NUMBER OF SEQ ID NOS: 44  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 1  
; LENGTH: 100  
; TYPE: PRT  
; ORGANISM: Bacillus sp. 1-1  
US-09-430-702-1  
Query Match 10.6%; Score 66.5; DB 4; Length 100;  
Best Local Similarity 25.9%; Pred. No. 1.1;  
Matches 15; Conservative 10; Mismatches 22; Indels 11; Gaps 1;  
QY 12 TAKHPWFQIEDRCYINDGKLFARGSIYGNMRFVDPKADYGGVGENLYHADVER 69  
DB 40 TPVHSSPALETNPVYNGAIVDNGALLGN-----YSNDQNLFFHNGTNP 86  
RESULT 11  
US-08-960-780-4  
; Sequence 4, Application US/08960780  
; Patent No. 6204435  
; GENERAL INFORMATION:  
; APPLICANT: Fetteison, Jerald S.  
; APPLICANT: Schmepl, H. Ernest  
; APPLICANT: Narva, Kenneth E.  
; APPLICANT: Stochhoff, Brian A.  
; APPLICANT: Schmeits, James  
; APPLICANT: Loewer, David  
; APPLICANT: Dullum, Charles Joseph  
; APPLICANT: Muller-Cohn, Judy  
; APPLICANT: Stamp, Lisa  
; TITLE OF INVENTION: No. 6204435el Pesticidal Toxins and Nucleotide  
; TITLE OF INVENTION: Sequences Which Encode These Toxins  
; NUMBER OF SEQUENCES: 134  
; CORRESPONDENCE ADDRESS: 134  
; ADDRESSEE: Saliwanchik, Lloyd & Saliwanchik  
; STREET: 2421 N.W. 41st Street, Suite A-1  
; CITY: Gainesville  
; STATE: FL  
; COUNTRY: US  
; ZIP: 32606-6669  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/960,780  
; FILING DATE: 30-OCT-1997  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/029,848  
; FILING DATE: 30-OCT-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Saliwanchik, David R.  
; REGISTRATION NUMBER: 31,794  
; REFERENCE/DOCKET NUMBER: MA-708  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 352-375-8100  
; TELEFAX: 352-372-5800  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 790 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; ORIGINAL SOURCE:  
; INDIVIDUAL ISOLATE: 364  
US-08-960-780-4  
Query Match 10.5%; Score 66; DB 3; Length 790;  
Best Local Similarity 23.4%; Pred. No. 21;







```

; GENERAL INFORMATION:
; APPLICANT: Crew, Mark D.
; TITLE OF INVENTION: Viral Stealth Technology to Prevent T-Cell-Mediated Rejection of
; FILE OF INVENTION: Xenografts
; FILE REFERENCE: 8793-50391
; CURRENT APPLICATION NUMBER: US/10/300,393
; PRIOR FILING DATE: 2002-11-19
; PRIOR APPLICATION NUMBER: US 60/342,981
; PRIOR FILING DATE: 2001-12-18
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: Patentin version 3.1
; LENGTH: 186
; SEQ ID NO 20
; TYPE: PRT
; ORGANISM: human cytomegalovirus Towne strain US3
US-10-300-393-20

Query Match          14.5%; Score 91; DB 15; Length 186;
Best Local Similarity 27.4%; Pred. No. 0.0069;
Matches 31; Conservative 20; Mismatches 38; Indels 24; Gaps 6;

QY 18 FOIEDNRCYIDNGKLPARGSIGVGNMSR-----FVFPDPRADYGGVGENLYVHADVEFV 70
DB 37 FRVEENQCSHMGMLHYKGMSSGNFTEKHFVSVGISQSYMDRLQVSGEQYHHDE----- 91
QY 71 PGSLSKNNV-----RNLDPVPIFETLRLRVLQGD-VIMLRCPVELRDVYTS 116
DB 92 RGAIFENNIGHFVPHVVDVNDI--TLSTR---WGDPKXYAACVPOVRMNYSS 139

RESULT 3
US-10-369-493-2939
; Sequence 2939, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 2939
; LENGTH: 301
; TYPE: PRT
; ORGANISM: Thermotoga maritima
US-10-369-493-2939

Query Match          12.0%; Score 75.5; DB 12; Length 301;
Best Local Similarity 32.3%; Pred. No. 0.96;
Matches 30; Conservative 11; Mismatches 33; Indels 19; Gaps 5;

QY 19 QIEDNRCYIDNGKLPARG-----SIVGNMSRFVFPDPRAD---YGVGENLYVHADVE 68
DB 205 RLSDHVCIIIDHGKIAGCTSSLSLSSGLKTVVFGDGDVNVAVYLERKENHYVETD-- 262
QY 69 FVPGESLK-----WNVNRLDV-WPIFETLRL 95
DB 263 -APEELIKDLKNNVNSNIVRKPNLEDFLKL 294

RESULT 4
US-10-014-436-4
; Sequence 4, Application US/10014436
; Publication No. US20020182699A1
; GENERAL INFORMATION:
; APPLICANT: HATADA, Yuji
```

```

; APPLICANT: IGARASHI, Kazuaki
; APPLICANT: OZAKI, Katsuya
; APPLICANT: ARA, Katsutoshi
; APPLICANT: KAWAI, Shuji
; APPLICANT: ITO, Susumu
; TITLE OF INVENTION: Gene for Enzyme Having Both Alkaline Pullulanase and Alkaline Alpi
; FILE REFERENCE: 2173-0122P
; CURRENT APPLICATION NUMBER: US/10/014,436
; CURRENT FILING DATE: 2002-05-21
; PRIOR APPLICATION NUMBER: JP 111547/1995
; PRIOR FILING DATE: 1995-05-10
; PRIOR APPLICATION NUMBER: PCT/JP96/01243
; PRIOR FILING DATE: 1996-05-10
; PRIOR APPLICATION NUMBER: US 08/952,084
; PRIOR FILING DATE: 1997-11-10
; PRIOR APPLICATION NUMBER: US 09/514,302
; PRIOR FILING DATE: 2000-02-28
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 2
; LENGTH: 1938
; TYPE: PRT
; ORGANISM: Bacillus sp.
US-10-014-436-2

Query Match          11.0%; Score 69; DB 14; Length 893;
Best Local Similarity 32.0%; Pred. No. 23;
Matches 24; Conservative 11; Mismatches 30; Indels 10; Gaps 3;

QY 12 TAKHPWFOIEDNRCYIDNGKLPARGSIGVGNMSRFVFPDPRAD-----YGVGENLYVHAD 65
DB 82 SVKSGWKLIDEMVAY--DGLGALHEDGTATLKVSPADNVSVLYXKVDQNEV--D 137
QY 66 DVEFVPGESLSKNNVR 80
DB 138 TIEMVKGDRGWSVK 152

RESULT 5
US-10-014-436-2
; Sequence 2, Application US/10014436
; Publication No. US20020182699A1
; GENERAL INFORMATION:
; APPLICANT: HATADA, Yuji
; APPLICANT: OZAKI, Katsuya
; APPLICANT: ARA, Katsutoshi
; APPLICANT: KAWAI, Shuji
; APPLICANT: ITO, Susumu
; TITLE OF INVENTION: Gene for Enzyme Having Both Alkaline Pullulanase and Alkaline Alpi
; FILE REFERENCE: 2173-0122P
; CURRENT APPLICATION NUMBER: US/10/014,436
; CURRENT FILING DATE: 2002-05-21
; PRIOR APPLICATION NUMBER: JP 111547/1995
; PRIOR FILING DATE: 1995-05-10
; PRIOR APPLICATION NUMBER: PCT/JP96/01243
; PRIOR FILING DATE: 1996-05-10
; PRIOR APPLICATION NUMBER: US 08/952,084
; PRIOR FILING DATE: 1997-11-10
; PRIOR APPLICATION NUMBER: US 09/514,302
; PRIOR FILING DATE: 2000-02-28
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 2
; LENGTH: 1938
; TYPE: PRT
; ORGANISM: Bacillus sp.
US-10-014-436-2

Query Match          11.0%; Score 69; DB 14; Length 1938;
Best Local Similarity 32.0%; Pred. No. 59;
```



	Matches	24;	Conservative	11;	Mismatches	30;	Indels	10;	Gaps	3;
QY	12	TAKHPWQIENRCYINDGKLFARGSTVGNMSRFVDEKAD-----YGGGENTLVYAD	65							
Db	1127	SVKSGWKLIDEMVY--DGKLGALHEHGTATLKWSKADNVSVVLVYDKYDQNEV--D	1182							
QY	66	DVEFVPGESLKNVYR	80							
Db	1183	TIEMWKDGRGVMSVK	1197							

```

RESULT 6
US-10-369-493-7954
; Sequence 7954, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; TITLE OF INVENTION: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(652052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 7954
; LENGTH: 384
; TYPE: PR1
; ORGANISM: Rhodobacter sphaeroides
; US-10-369-493-7954

```

[illegible]

```

RESULT 7
US-10-375-932-25
Sequence 25, Application US/10375932
Publication No. US20040009469A1
GENERAL INFORMATION:
APPLICANT: Apt, Doris
APPLICANT: Punnonen, Juha
APPLICANT: Brinkman, Alice M.
TITLE OF INVENTION: NOVEL FLAVIVIRUS ANTIGENS
FILE REFERENCE: 0322.210US
CURRENT APPLICATION NUMBER: US/10/375,932
CURRENT FILING DATE: 2003-07-26
PRIOR APPLICATION NUMBER: US 60/360,030
PRIOR FILING DATE: 2002-02-26
NUMBER OF SEQ ID NOS: 345
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO 25
LENGTH: 448
TYPE: PR1
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Synthetic polypeptide: 16B4-NPRM
US-10-375-932-25

```

Query Match 10.7%; Score 67.5; DB 12; Length 448;

[illegible]

```

RESULT 8
US-10-375-932-89
Sequence 89, Application US/103759932
Publication No. US20040009469A1
GENERAL INFORMATION:
APPLICANT: Apt, Doris
APPLICANT: Punnonen, Juhana
APPLICANT: Brinkman, Alice M.
TITLE OF INVENTION: NOVEL FLAVIVIRUS ANTIGENS
FILE REFERENCE: 0322.210US
CURRENT APPLICATION NUMBER: US/10/375,932
CURRENT FILING DATE: 2003-02-26
PRIORITY FILING DATE: 2002-02-26
PRIORITY FILING DATE: 2002-02-26
NUMBER OF SEQ ID NOS: 345
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 89
LENGTH: 464
TYPE: PR1
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Synthetic polypeptide: 16B4
US-10-375-932-89

```

	Query Match	Similarity	25.0%	Score 67.5	DB 12	Length 464
	Best Local	Matches	Conservative	28	Mismatches	43
					Indels	29
					Gaps	6
Oy	2	GEDALRPWKSTAPKPMFQIEDNR	CYINDGKLFARGSI	GNMRFV-FDPKADYG-----	54	
Db	343	GEDD-----PCKLP-FSTDEK	GVTONGRILITANPI	YTEKDSFNIDAEPPFGS	IYV	395
Oy	55	GVGNELVYHADDFEVFGESL	KNNVRNLDWPIETL---	ALRLVQSDVIW	103	
Db	396	GVGDKAL-----KINWYK	KGSSISQMFETTTIR	GAKRAILGDTAW	435	

```

RESULT 9
: US-10-369-493-585
: Sequence 585, Application US/10369493
: Publication No. US2003023675A1
: GENERAL INFORMATION:
: APPLICANT: Cao, Yongwei
: APPLICANT: Hinkle, Gregory J.
: APPLICANT: Slater, Steven C.
: APPLICANT: Goldman, Barry S.
: APPLICANT: Chen, Xianfeng
: TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
: TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
: FILE REFERENCE: 38-10(52052)B
: CURRENT APPLICATION NUMBER: US/10/369,493
: PRIOR FILING DATE: 2003-02-28
: PRIOR APPLICATION NUMBER: US 60/360,039
: PRIOR FILING DATE: 2002-02-21
: NUMBER OF SEQ ID NOS: 47374
: SEQ ID NO 585
: LENGTH: 564
: TYPE: PRT
: ORGANISM: Deinococcus radiodurans
: US-10-369-493-585

```

```

Query Match Similarity    10.7%; Score 67.5; DB 12; Length 6291;
Best Local Similarity     30.0%; Pred. No. 3.9e+02;
Matches   24; Conservative    7; Mismatches   22; Indels    27; Gaps    3;

Cy      3 EDALRPWKSTAKHPWFQIEDNRCYIDNGKLPARGSIVG-----NMSEFVDPADRY 53
       : ||||| |-----TGELYLGRGRLARGLRRPGLTAARFVAD---H 822
       : ||||| |-----
Db      781 DGLAPRVPDGA-----
Cy      54 GGVEGNLYIHADDVFEPVGE 73
       : ||| : | ||||
Db      823 TGTGERMYRTGLVRRVPE 842

RESULT 11
US-09-853-161-100
; Sequence 100, Application US/09853161
; Patent No. US20020076756A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: 28 Human Secreted Proteins
FILE REFERENCE: P2003P3
CURRENT APPLICATION NUMBER: US/09/853,161
CURRENT FILING DATE: 2001-05-11
PRIOR APPLICATION NUMBER: 60/265,583
PRIOR FILING DATE: 2001-02-02
PRIOR APPLICATION NUMBER: 09/152,060
PRIOR FILING DATE: 1998-09-11
PRIOR APPLICATION NUMBER: PCT/US98/04858
PRIOR FILING DATE: 1998-03-12
PRIOR APPLICATION NUMBER: 60/040,762
PRIOR FILING DATE: 1997-03-14
PRIOR APPLICATION NUMBER: 60/040,710
PRIOR FILING DATE: 1997-03-14
PRIOR APPLICATION NUMBER: 60/050,934
PRIOR FILING DATE: 1997-05-30
PRIOR APPLICATION NUMBER: 60/048,100

```

```

RESULT 12
US-09-852-659A-100
; Sequence 100, Application US/09852659A
; Patent No. US2002007287A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 28 Human Secreted Proteins
; FILE REFERENCE: P2003P4
; CURRENT FILING DATE: 2001-05-11
; PRIOR APPLICATION NUMBER: 60/265,583
; PRIOR FILING DATE: 2001-02-02
; PRIOR APPLICATION NUMBER: 09/152,060
; PRIOR FILING DATE: 1998-09-11
; PRIOR APPLICATION NUMBER: PCT/US98/04858
; PRIOR FILING DATE: 1998-03-12
; PRIOR APPLICATION NUMBER: 60/040,762
; PRIOR FILING DATE: 1997-03-14
; PRIOR APPLICATION NUMBER: 60/040,710
; PRIOR FILING DATE: 1997-03-14
; PRIOR APPLICATION NUMBER: 60/050,934
; PRIOR FILING DATE: 1997-05-30
; PRIOR APPLICATION NUMBER: 60/048,100
; PRIOR FILING DATE: 1997-05-30
; PRIOR APPLICATION NUMBER: 60/048,357
; PRIOR FILING DATE: 1997-05-30
; PRIOR APPLICATION NUMBER: 60/048,189
; PRIOR FILING DATE: 1997-05-30
; PRIOR APPLICATION NUMBER: 60/057,765
; PRIOR FILING DATE: 1997-09-05
; PRIOR APPLICATION NUMBER: 60/048,970
; PRIOR FILING DATE: 1997-06-06
; PRIOR APPLICATION NUMBER: 60/068,368
; PRIOR FILING DATE: 1997-12-19
; NUMBER OF SEQ ID NOS: 121
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 100
; LENGTH: 240
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-852-659A-100

```

[illegible][illegible]

```
; PRIOR FILING DATE: 1998-03-12
; PRIOR APPLICATION NUMBER: 60/040,762
; PRIOR FILING DATE: 1997-03-14
; PRIOR APPLICATION NUMBER: 60/040,710
; PRIOR FILING DATE: 1997-03-14
; PRIOR APPLICATION NUMBER: 60/050,934
; PRIOR FILING DATE: 1997-05-30
; PRIOR APPLICATION NUMBER: 60/048,100
; PRIOR FILING DATE: 1997-05-30
; PRIOR APPLICATION NUMBER: 60/048,357
; PRIOR FILING DATE: 1997-05-30
; PRIOR APPLICATION NUMBER: 60/048,189
; PRIOR FILING DATE: 1997-05-30
; PRIOR APPLICATION NUMBER: 60/057,765
; PRIOR FILING DATE: 1997-09-05
; PRIOR APPLICATION NUMBER: 60/048,970
; PRIOR FILING DATE: 1997-06-06
; PRIOR APPLICATION NUMBER: 60/068,368
; PRIOR FILING DATE: 1997-12-19
; NUMBER OF SEQ ID NOS: 121
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 64
; LENGTH: 335
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (35)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (297)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-852-659A-64

Query Match      10.7%; Score 67; DB 9; Length 335;
Best Local Similarity 26.4%; Pred. No. 12;
Matches 29; Conservative 18; Mismatches 35; Indels 28; Gaps 6;

QY 14 KHPFQIED-----NRCYIDNGKLPARGSIYGNMSRFVDPKADYGVGEN----- 59
DB 151 RHAYVQAENHLQTAAGTCGVSDSL--GSLGPRTAAVFRPRP-----GDSLPSRETRYV 202

QY 60 -LYVHADDFE-VPGESLKMNVRNLDVM---PIFETLALRLVLQGDVIW 103
DB 203 ELIYVVVNAERQMGSEAAVNRHRYLEVYVNNHVDKLYOKLNRVVVLVGLIIN 252
```

Search completed: January 22, 2004, 12:09:28  
Job time : 283.676 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 22, 2004, 11:37:25 ; Search time 21.3788 Seconds  
(without alignments)  
521.804 Million cell updates/sec

Title: US-09-890-806-3\_COPY\_28\_143

Sequence: 1 AGEDALRPWKSTAKHPWFQI.....LQGVIMLRVPELRVDYTS 116

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database :  
1: p1r1:\*  
2: p1r2:\*  
3: p1r3:\*  
4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	639	100.0	199	1 QOBECS	HQLF2 protein prec
2	97	15.4	186	1 QOBECS	HQLF1 precursor -
3	78.5	12.5	254	2 AD0205	probable N-acetyl
4	78.5	12.5	415	2 T21532	hypothetical prote
5	77.5	12.3	703	1 ALBEX1	cyclomaltodextrin
6	75.5	12.0	301	2 E72384	ABC transporter, A
7	74	11.8	261	2 A98850	hypothetical prote
8	74	11.8	261	2 E85798	hypothetical prote
9	74	11.8	261	2 C64948	probable membrane
10	74	11.8	360	2 AI2854	conserved hypothet
11	74	11.8	360	2 E97631	hypothetical prote
12	72.5	11.5	159	1 E71208	hypothetical prote
13	72	11.4	1409	1 OFFFCP	copa polypeptide
14	71	11.3	2201	2 AH0095	probable sideropho
15	70.5	11.2	394	2 AF2963	phosphomannose iso
16	70.5	11.2	704	2 G98319	mannose-6-phosphat
17	70.5	11.2	704	2 I39805	cyclomaltodextrin
18	70	11.1	271	2 C72307	hypothetical prote
19	70	11.1	588	2 T48009	PECTINESTERASE-1Lk
20	69.5	11.0	349	1 QOBECS	HRRF protein - hu
21	68.5	10.9	273	2 D81678	probable oxidoredu
22	68.5	10.9	314	2 D97538	ribosomal large ch
23	68.5	10.9	329	2 AG2757	hypothetical prote
24	68.5	10.9	347	2 D83605	probable binding p
25	68.5	10.9	359	2 A10140	molybdenum transpo
26	68.5	10.9	1629	2 T06461	DNA-binding protei
27	68	10.8	689	2 F81286	probable polysacch
28	68	10.8	1554	2 C72647	hypothetical prote
29	67.5	10.7	564	2 H75403	glycosyl hydrolase

30	67	10.7	367	2 T09376	hypothetical prote
31	67	10.7	381	2 S69795	UTP-hexose-1-phosp
32	67	10.7	757	2 B90572	lipoprotein (impor
33	67	10.7	1690	2 T40847	probable RNA biog
34	67	10.7	7463	2 T36248	CDA peptide synthet
35	66.5	10.6	159	2 F75021	ribosomal protein
36	66.5	10.6	233	2 E86340	protein F2D10.32 (
37	66.5	10.6	463	2 D41853	hexose phosphate t
38	66.5	10.6	463	2 AC0964	hexosephosphate tr
39	66.5	10.6	1151	2 S48431	probable membrane
40	66.5	10.6	1208	2 T27822	hypothetical prote
41	66	10.5	261	2 A10742	high-affinity zinc
42	66	10.5	361	2 JN0716	glutamate-ammonia
43	65.5	10.4	273	2 H81140	oxidoreductase, sh
44	65.5	10.4	286	2 T38680	probable fo-to par
45	65.5	10.4	489	2 G91143	ribonuclease G [im

## ALIGNMENTS

## RESULT 1

QOBECS HQLF2 protein precursor - human cytomegalovirus (strain AD169)

N:Alternate names: hypothetical protein US2

C:Species: human cytomegalovirus, human herpesvirus 5

C>Date: 30-Sep-1989 #sequence\_revision 30-Sep-1989 #text\_change 28-Jul-2000

C/Accession: E26078, S09916

R:Weston, K.; Bartell, B.G.

J. Mol. Biol. 192, 177-208 (1986)

A>Title: Sequence of the short unique region, short repeats, and part of the long repeats

A:Reference number: A92935; MUID:87169717; PMID:3031311

A/Accession: E26078

A:Molecule type: DNA

A:Residues: 1-199 <CH>

A/Cross-references: EMBL:X04650; NID:G59801; PIDN:CAH37096.1; PID:G4456177

A/Experimental source: strain AD169

R:Chen, M.S.; Bankier, A.T.; Beck, S.; Bohm, R.; Brown, C.M.; Cerny, R.; Hornsall, T.; J

M.; Bartell, B.G.

Curr. Top. Microbiol. Immunol. 154, 125-169, (1989)

A>Title: Analysis of the protein-coding content of the sequence of human cytomegalovirus

A:Reference number: S09749; MUID:90269039; PMID:1613119

A/Accession: S09916

A>Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-199 <CH>

A/Cross-references: EMBL:X17403; NID:G59591; PIDN:CAH35313.1; PID:G1780933

A/Experimental source: strain AD169

A/Note: this sequence was submitted to the EMBL Data Library, December 1989

C:Genetics:

A:Gene: HQLF2

C:Superfamily: cytomegalovirus HQLF2 protein

C:Keywords: glycoprotein; transmembrane protein

F:1-25/Domain: signal sequence #status predicted <SIG>

F:26-199/Product: hypothetical protein US2 #status predicted <MAT>

F:161-186/Domain: transmembrane #status predicted <TM>

F:68/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 100.0%; Score 629; DB 1; Length 199;  
Best Local Similarity 100.0%; Pred. No. 5.8e-58;  
Matches 116; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	AGEDALRPWKSTAKHPWFQIEDNRCYINGKLPARGSTVGNMSPFDPKAYGGGENTL	60
DB	28	AGEDALRPWKSTAKHPWFQIEDNRCYINGKLPARGSTVGNMSPFDPKAYGGGENTL	87
QY	61	YHADDVEFPGESLKMVRNDVMPFETTLALRLVLQGVIMLRVPELRVDYTS	116
DB	88	YHADDVEFPGESLKMVRNDVMPFETTLALRLVLQGVIMLRVPELRVDYTS	143

RESULT 2  
QOBECS

HOLFI precursor - human cytomegalovirus (strain AD169)  
 N/Alternate names: hypothetical protein US3  
 C/Species: human cytomegalovirus, human herpesvirus 5  
 C/Date: 30-Sep-1989 #sequence\_revision 30-Sep-1989 #text\_change 28-Jul-2000  
 C/Accession: F26078; S09917  
 R/Watson, K.; Barrell, B.G.  
 J. Mol. Biol. 192, 177-208, 1986  
 A/Title: Sequence of the short unique region, short repeats, and part of the long repeat  
 A/Reference number: A92935; MUID:87169717; PMID:3031311  
 A/Accession: F26078  
 A/Molecule type: DNA  
 A/Residues: 1-186 <MES>  
 A/Cross-references: EMBL:X04650; NID:G59801; PIDN:CA837097.1; PID:G4456178  
 A/Experimental source: strain AD169  
 R/Chee, M.S.; Bankier, A.T.; Beck, S.; Bohm, R.; Brown, C.M.; Cerny, R.; Horsnall, T.; M.; Barrell, B.G.  
 Curr. Top. Microbiol. Immunol. 154, 125-169, 1990  
 A/Title: Analysis of the protein-coding content of the sequence of human cytomegalovirus  
 A/Reference number: S09749; MUID:90269039; PMID:2161319  
 A/Accession: S09917  
 A/Status: nucleic acid sequence not shown; translation not shown  
 A/Molecule type: DNA  
 A/Residues: 1-186 <CHE>  
 A/Cross-references: EMBL:X17403; NID:G59591; PIDN:CA835314.1; PID:G1780934  
 A/Experimental source: strain AD169  
 C/Genetics:  
 A/Note: this sequence was submitted to the EMBL Data Library, December 1989  
 A/Genes: HOLFI  
 C/Superfamily: cytomegalovirus HOLFI protein  
 C/Keywords: glycoprotein; transmembrane protein  
 F/1-25/Domain: signal sequence #status predicted <SIG>  
 F/26-186/Product: hypothetical protein US3 #status predicted <MAT>  
 F/160-182/Domain: transmembrane #status predicted <TM>  
 F/60/Binding site: carbohydrate (asn) (covalent) #status predicted

Query Match 15.4%; Score 97; DB 1; Length 186;  
 Best Local Similarity 27.0%; Pred. No. 0.01;  
 Matches 31; Conservative 24; Mismatches 32; Indels 28; Gaps 7;

QY 18 FQIEDNRCYIDNGKLFARGSIIVGMSRFVDPKADYGVGENTL-----YVHADVE 68  
 DB 37 FRYEENQCFHMGMLTYFKGMSGNFTEKHF---VNVGIVGQSYMDRLQVSGEYVHDE-- 91

QY 69 FVGESESLKNNV-----RNDVMPFETLALRLVLOGD-VYIMRCVPELVDTYS 116  
 DB 92 --RGATFEMNIGGHPTHVDMVDI--TLSTR--WGPCKKIACVQVMDISS 139

RESULT 3  
 AD0205  
 Probable N-acetylmuramoyl-L-alanine amidase (imported) - Yersinia pestis (strain CO92)  
 C/Species: Yersinia pestis  
 C/Date: 02-Nov-2001 #sequence\_revision 02-Nov-2001 #text\_change 02-Nov-2001  
 C/Accession: AD0205  
 R/Parhill, J.; Wren, B.W.; Thomson, N.R.; Tiltball, R.W.; Holden, M.T.G.; Prentice, M.B.; deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davis, P.; Dougan, G.; H., M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell, Nature 413, 523-527, 2001  
 A/Title: Genome sequence of Yersinia pestis, the causative agent of plague.  
 A/Reference number: AB0001; MUID:21470413; PMID:11586360  
 A/Accession: AD0205  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-254 <KUR>  
 A/Cross-references: GB:AL590842; PIDN:CAC90504.1; PID:G15979715; GSPDB:GN00175  
 C/Genetics:  
 A/Genes: YP01683

Query Match 12.5%; Score 78.5; DB 2; Length 254;  
 Best Local Similarity 25.3%; Pred. No. 1.2;  
 Matches 24; Conservative 15; Mismatches 43; Indels 13; Gaps 3;

QY 8 PWKS---TAKHPWQIEDNRCYIDN--GKLFARSGIIVGMSRFVDPKADYGVGENTLYV 62

DB 162 PKALYDAGIGAMYDDETKQRYLDQFLCSLPSKNNIISLTKRYGVDTSAGVSEVGNLI 221  
 QY 63 HADDEVFVGESESLKNNVRRLDVMPIFETLALV 97  
 DB 222 RAFLHFRP-----CNVDGIPDAETVALYAL 248

RESULT 4  
 T21532  
 Hypothetical protein F28H7.7 - Caenorhabditis elegans  
 C/Species: Caenorhabditis elegans  
 C/Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
 C/Accession: T21532  
 R/Berks, M.  
 submitted to the EMBL Data Library, May 1996  
 A/Reference number: Z19435  
 A/Accession: T21532  
 A/Status: preliminary; translated from GB/EMBL/DBJ  
 A/Molecule type: DNA  
 A/Residues: 1-415 <WIL>  
 A/Cross-references: EMBL:Z72508; PIDN:CA96643.1; GSPDB:GN00023; CESP:F28H7.7  
 A/Experimental source: clone F28H7  
 C/Genetics:  
 A/Genes: CESP:F28H7.7  
 A/Map position: 5  
 A/Introns: 83/2; 147/1; 180/3; 231/3; 299/3; 364/2

Query Match 12.5%; Score 78.5; DB 2; Length 415;  
 Best Local Similarity 29.2%; Pred. No. 2.2;  
 Matches 26; Conservative 9; Mismatches 31; Indels 23; Gaps 6;

QY 7 RPKSKTAKHPWQIEDNRCY-----IDNGKLFAR-----GSIVGMSRFVDPKADYGV 56  
 DB 307 RKMELTTO--PMLNCTDGGKCFSGMAYNNQVLCQMLHDVGEVNGVYQTLIDPKICY--- 362

QY 57 GENTLYHADDEVFVGESESLKNNVRRLDVM 85  
 DB 363 ----IRHYRDVMMG-----KMMVRNNDVL 382

RESULT 5  
 ALBXX1  
 Cyclodextrin glucanotransferase (EC 2.4.1.19) precursor - Bacillus sp. (strain 1-1)  
 N/Alternate names: cyclodextrin glycosyltransferase  
 C/Species: Bacillus sp.  
 C/Date: 31-Mar-1993 #sequence\_revision 31-Mar-1993 #text\_change 24-Apr-1998  
 C/Accession: S26399; S26593  
 R/Schmid, G.; Engelbrecht, A.; Schmid, D.  
 in Proceedings of the Fourth International Symposium on Cyclodextrins, Huber, O., and Sz  
 A/Title: Cloning and nucleotide sequence of a cyclodextrin glycosyltransferase gene from  
 A/Reference number: S26399  
 A/Accession: S26399  
 A/Status: not compared with conceptual translation  
 A/Molecule type: DNA  
 A/Residues: 1-703 <SCH>  
 A/Accession: S26593  
 A/Molecule type: protein  
 A/Residues: 30-52 <SC2>  
 C/Function:  
 A/Description: catalyzes the breaking and reformation of 1,4-alpha-D-glucopyranosyl bonds  
 A/Note: can also catalyze the disproportionation transfer of 6 to 8 1,4-alpha-D-glucopyr  
 C/Superfamily: cyclomaltoextrin glucanotransferase; alpha-amylase core homology  
 C/Keywords: calcium; duplication; extracellular protein; glycosyltransferase; hexosyltrai  
 F/1-29/Domain: signal sequence #status predicted <SIG>  
 F/30-703/Product: cyclomaltoextrin glucanotransferase #status experimental <MAT>  
 F/161-224/Domain: B <DOB>  
 F/30-160/Domain: A1 <DA1>  
 F/219-353/Domain: alpha-amylase core homology <AMY>  
 F/225-428/Domain: A2 <DA2>  
 F/429-516/Domain: C <DOC>  
 F/517-600/Domain: D <DOD>  
 F/601-703/Domain: E <DOE>

F:52,54,57,59,76,78/Binding site: calcium (Asp, Asn, Asn, Gly, Asp) #status predicted  
F:68-75/Disulfide bonds: #status predicted  
F:161,212,221,255/Binding site: calcium (Asn, Ile, Asp, His) #status predicted  
F:251,279,350/Active site: Asp, Glu, Asp #status predicted

Query Match 12.3%; Score 77.5; DB 1; Length 703;  
Best Local Similarity 24.0%; Pred. No. 5.3;

Matches 25; Conservative 22; Mismatches 44; Indels 13; Gaps 3;

12 TAKHPFOJEDNRCYIDNGKLFARGSIYGMRSFVDPKADYGVGBENLYVHADVDFVP 71

159 TPNHSSPALETNNYVNGAIYDNGALNGN-----YSDQONLFHNHGTDFSS 207

72 GE-SLKVNRYLDVMPFIETLRLVYQGDVWL-RCVBELRYD 113

208 YEDSIYRNLVDLADYDNLNTVMQYKESIKFWLDGIDGIRVD 251

RESULT 6

ABC transporter, ATP-binding protein - Thermotoga maritima (strain MSB8)

C:Species: Thermotoga maritima  
C:Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 21-Jul-2000

C:Accession: E72384  
R:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickey

Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.;  
C.M.

Nature 399, 323-329, 1999  
A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome seq

A:Reference number: A72200; WUID:99287316; PMID:10360571

A:Accession: E72384

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-301 <ARN>

A:Cross-references: GB:AE001718; GB:AE00512; NID:94980881; PIDN:AAJ35474.1; PID:9498089

A:Experimental source: strain MSB8

C:Genetics:

A:Gene: TW0389

C:Superfamily: unassigned ATP-binding cassette proteins; ATP-binding cassette homology

Query Match 12.0%; Score 75.5; DB 2; Length 301;  
Best Local Similarity 32.3%; Pred. No. 3.1;

Matches 30; Conservative 11; Mismatches 33; Indels 19; Gaps 5;

19 QIEDNRCYIDNGKLFARG-----STVGNMRSFVDPKAD---YGVGENLYVHADV 68

205 RUSDHVCITIDHGKIAGEFPSSLISSGKTVAFPCDDQVNVRYLEKKNHVVETD-- 262

69 FVPGESLK-----MNVRLDV-MPIETLRL 95

263 -APEELIKDLKKNVSNVIRKNEIDVFLK 294

RESULT 7

A98950  
hypothetical protein ECG2569 [imported] - Escherichia coli (strain O157:H7, substrain R1

C:Species: Escherichia coli  
C:Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 03-Aug-2001

C:Accession: A98950

R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.

gaasawa, N.; Yasunaga, T.; Kohara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.

DNA Res. 8, 11-22, 2001  
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gene

A:Reference number: A98629; WUID:21156231; PMID:11258796

Query Match 11.8%; Score 74; DB 2; Length 261;  
Best Local Similarity 40.6%; Pred. No. 3.8;

Matches 13; Conservative 8; Mismatches 11; Indels 0; Gaps 0;

82 LDVMPFIETLRLVYQGDVWLRCVBELRYD 113

52 LDVNPYAVIATVTLLAGGLVWLKRPOLAD 83

RESULT 8

E85798  
hypothetical protein yeb1 [imported] - Escherichia coli (strain O157:H7, substrain EDL933

C:Species: Escherichia coli  
C:Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 14-Sep-2001

C:Accession: E85798

R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glaesner, U.D.; Rose, D.J.; Mayhew,

Iller, L.; Grobeck, E.J.; Davis, N.W.; Lam, A.; Dimalanta, E.; Potamousis, K.; Apodaca,

Nature 409, 529-533, 2001

A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.

A:Reference number: A85480; WUID:21074935; PMID:11206551

A:Accession: E85798

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-261 <STO>

A:Cross-references: GB:AE005174; NID:912515915; PIDN:AAJ35474.1; GSPDB:GN00145; UMGF:Z293

A:Experimental source: strain O157:H7, substrain EDL933

C:Genetics:

A:Gene: yeb1

C:Superfamily: conserved hypothetical protein H10360

Query Match 11.8%; Score 74; DB 2; Length 261;  
Best Local Similarity 40.6%; Pred. No. 3.8;

Matches 13; Conservative 8; Mismatches 11; Indels 0; Gaps 0;

82 LDVMPFIETLRLVYQGDVWLRCVBELRYD 113

52 LDVNPYAVIATVTLLAGGLVWLKRPOLAD 83

RESULT 9

C64948  
probable membrane protein yeb1 - Escherichia coli (strain K-12)

C:Species: Escherichia coli  
C:Date: 12-Sep-1997 #sequence\_revision 17-Sep-1997 #text\_change 01-Mar-2002

C:Accession: C64948

R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co

Science 277, 1453-1462, 1997

A:Title: The complete genome sequence of Escherichia coli K-12.

A:Reference number: A64720; WUID:97426617; PMID:9278503

A:Accession: C64948

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-261 <BLAT>

A:Cross-references: GB:AB000280; GB:U00096; NID:91788163; PIDN:AACT4929.1; PID:91788166;

A:Experimental source: strain K-12, substrain MG1655

C:Genetics:

A:Gene: yeb1

C:Superfamily: conserved hypothetical protein H10360

Query Match 11.8%; Score 74; DB 2; Length 261;  
Best Local Similarity 40.6%; Pred. No. 3.8;

Matches 13; Conservative 8; Mismatches 11; Indels 0; Gaps 0;

82 LDVMPFIETLRLVYQGDVWLRCVBELRYD 113

52 LDVNPYAVIATVTLLAGGLVWLKRPOLAD 83

RESULT 9

C64948  
probable membrane protein yeb1 - Escherichia coli (strain K-12)

C:Species: Escherichia coli  
C:Date: 12-Sep-1997 #sequence\_revision 17-Sep-1997 #text\_change 01-Mar-2002

C:Accession: C64948

R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co

Science 277, 1453-1462, 1997

A:Title: The complete genome sequence of Escherichia coli K-12.

A:Reference number: A64720; WUID:97426617; PMID:9278503

A:Accession: C64948

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-261 <BLAT>

A:Cross-references: GB:AB000280; GB:U00096; NID:91788163; PIDN:AACT4929.1; PID:91788166;

A:Experimental source: strain K-12, substrain MG1655

C:Genetics:

A:Gene: yeb1

C:Superfamily: conserved hypothetical protein H10360

C:Keyword: transmembrane protein; transport protein

F:11-77/Domain: transmembrane #status predicted <TM1>

F:18-54/Domain: transmembrane #status predicted <TM2>

F:58-74/Domain: transmembrane #status predicted <TM3>

F:85-101/Domain: transmembrane #status predicted <TM4>

F:127-143/Domain: transmembrane #status predicted <TM5>

F:182-198/Domain: transmembrane #status predicted <TM6>

F:213-229/Domain: transmembrane #status predicted <TM7>

F:240-256/Domain: transmembrane #status predicted <TM8>

Query Match 11.8%; Score 74; DB 2; Length 261;  
Best Local Similarity 40.6%; Pred. No. 3.8;

Matches 13; Conservative 8; Mismatches 11; Indels 0; Gaps 0;

QY 82 LDVMPFETLRLVQGVIMLRCPPELRVD 113  
DB 52 LDVMPFYAVIATVLLAGLWLEKRPOLAI 83

## RESULT 10

AI2854  
conserved hypothetical protein Atu2266 [imported] - Agrobacterium tumefaciens (strain C5  
C/Species: Agrobacterium tumefaciens  
C/Date: 11-Jan-2002 #sequence\_revision 11-Jan-2002 #text\_change 18-Nov-2002  
C/Accession: AI2854  
R/Wood, D.W.; Seubal, J.C.; Kaul, R.; Monke, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I.  
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; Mclell  
Science 294, 2317-2323, 2001  
A/Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,  
ster, E.W.  
A/Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.  
A/Reference number: AB2577; MUID:21608550; PMID:11743193  
A/Accession: AI2854  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-360 <KUR>  
A/Cross-references: GB:AE008688; PIDN:AL43255.1; PID:G17740741; GSPDB:GN00186  
A/Experimental source: strain C58 (Dupont)  
C/Genetics:  
A/Gene: Atu2266  
A/Map position: circular chromosome  
C/Superfamily: Synchocystis hypothetical protein slr1087

Query Match 11.8%; Score 74; DB 2; Length 360;  
Best Local Similarity 26.6%; Pred. No. 5.5;  
Matches 17; Conservative 12; Mismatches 21; Indels 14; Gaps 2;

QY 10 KSTAKHFWFOIEDNRCYIDNGKLFARGSIYGNMRFVDPKADYGVGENLYVHADVER 69  
DB 295 RAVSKTWFMEINEDFGID-----IDNIESLAFGVVD---GKQIFVIASDNNF 340

QY 70 VPGE 73  
DB 341 NPGK 344

## RESULT 11

F97631  
hypothetical protein AGR\_C\_4118 [imported] - Agrobacterium tumefaciens (strain C58, Cere  
C/Species: Agrobacterium tumefaciens  
C/Date: 30-Sep-2001 #sequence\_revision 30-Sep-2001 #text\_change 18-Nov-2002  
C/Accession: F97631  
R/Godner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,  
A.; Liu, F.; Wollam, C.; Allinger, M.; Dougherty, D.; Scott, C.; Lappas, C.; Markelz, B.;  
Science 294, 2323-2328, 2001  
A/Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum  
A/Reference number: A97359; MUID:21608551; PMID:11743194  
A/Accession: F97631  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-360 <KUR>  
A/Cross-references: GB:AE007869; PIDN:AKX8007.1; PID:G15157422; GSPDB:GN00169  
C/Genetics:  
A/Gene: AGR\_C\_4118  
A/Map position: circular chromosome  
C/Superfamily: Synchocystis hypothetical protein slr1087

Query Match 11.8%; Score 74; DB 2; Length 360;  
Best Local Similarity 26.6%; Pred. No. 5.5;  
Matches 17; Conservative 12; Mismatches 21; Indels 14; Gaps 2;

QY 10 KSTAKHFWFOIEDNRCYIDNGKLFARGSIYGNMRFVDPKADYGVGENLYVHADVER 69  
DB 295 RAVSKTWFMEINEDFGID-----IDNIESLAFGVVD---GKQIFVIASDNNF 340

QY 70 VPGE 73  
DB 341 NPGK 344

## RESULT 12

E71208  
hypothetical protein PH1933 - Pyrococcus horikoshii  
C/Species: Pyrococcus horikoshii  
C/Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 21-Jul-2000  
C/Accession: E71208  
R/Kawabuchi, Y.; Sawada, M.; Horikawa, H.; Halkawa, Y.; Hino, Y.; Yamamoto, S.; Sekine  
M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushiida, N.; Oguchi,  
DNA Res. 5, 55-76, 1998  
A/Title: Complete sequence and gene organization of the genome of a hyper-thermophilic ar  
A/Reference number: A71000; MUID:98344137; PMID:9679194  
A/Accession: E71208  
A/Status: preliminary; nucleic acid sequence not shown; translation not shown  
A/Molecule type: DNA  
A/Residues: 1-159 <KAW>  
A/Cross-references: GB:AP000007; NID:G3236134; PIDN:BAA31060.1; PID:G3258377  
A/Experimental source: strain OT3  
A/Note: this accession replaces an interim accession for a sequence replaced by GenBank  
C/Genetics:  
A/Gene: PH1933  
C/Superfamily: Methanococcus jannaschii conserved hypothetical protein MJ1207

Query Match 11.5%; Score 72.5; DB 1; Length 159;  
Best Local Similarity 27.5%; Pred. No. 3;  
Matches 22; Conservative 12; Mismatches 31; Indels 15; Gaps 3;

QY 36 GSIVGNMRFVDPKADYGVGENLYV-----HADVERVPGSLKMNVRNLDVMP 87  
DB 82 GRVGAHFEVVDKFRQGGIGKRLITCLDFLGKXNDIIELVGB-----KNYGMNL 135

QY 88 FETLRLVQLQGVIMLRVCV 107  
DB 136 YEKGFVKVKGSG-IWVRMI 154

## RESULT 13

OFFPCP  
copia polyprotein - fruit fly (Drosophila melanogaster) retrotransposon copia  
N/Contains: copia protein, 31k; copia protein, 48k; proteinase  
C/Species: Drosophila melanogaster  
C/Date: 04-Dec-1986 #sequence\_revision 04-Dec-1986 #text\_change 16-Jul-1999  
C/Accession: A03324; S03612; S14835  
R/Mount, S.M.; Rubin, G.M.  
Mol. Cell. Biol. 5, 1630-1638, 1985  
A/Title: Complete nucleotide sequence of the Drosophila transposable element copia: homo)  
A/Reference number: A03324; MUID:85267679; PMID:2410772  
A/Accession: A03324  
A/Molecule type: DNA  
A/Residues: 1-1409 <MOU>  
A/Cross-references: GB:M1240; NID:G158615; PIDN:AA74497.1; PID:G950318  
R/Miller, K.; Rosenbaum, J.; Zbrzezna, V.; Pogo, A.O.  
Nucleic Acids Res. 17, 2131, 1989  
A/Title: The nucleotide sequence of Drosophila melanogaster copia-specific 2.1-kb mRNA.  
A/Reference number: S03612; MUID:89183629; PMID:2538606  
A/Accession: S03612  
A/Molecule type: mRNA  
A/Residues: 1-391,1375-1409 <ML>  
A/Cross-references: EMBL:X13719; NID:G7745; PIDN:CA3197.1; PID:G7746  
R/Yoshioke, K.; Honma, H.; Zushi, M.; Kondo, S.; Togashi, S.; Miyake, T.; Shiba, T.  
EMBO J. 9, 535-541, 1990  
A/Title: Virus-like particle formation of Drosophila copia through autocatalytic processes)  
A/Reference number: S14835; MUID:90151630; PMID:1689241  
A/Accession: S14835  
A/Molecule type: DNA  
A/Residues: 1-391,1375-1409 <YOS>  
A/Cross-references: EMBL:X54147; NID:G7749; PIDN:CA38086.1; PID:G7750  
C/Genetics:  
A/Gene: FlyBase:copia



A:Cross-references: F1/Baase:Pfgn0000349  
A:Mobile element: retrotransposon copia  
C:Superfamily: retrovirus-related polypotein  
C:Keywords: polypotein, proteinase  
F:2-473/Product: copia protein, 48k #status predicted <MAT1>  
F:2-270/Product: copia protein, 31k #status predicted <MAT2>  
F:271-453/Product: proteinase #status predicted <MAT>

Query Match 11.4%; Score 72; DB 1; Length 1409;  
Best Local Similarity 23.5%; Pred. No. 45;  
Matches 24; Conservative 18; Mismatches 34; Indels 26; Gaps 5;

```
Oy      17WFOIED---NRCYIDNGKUFARGSVIGVNNSRFVEDPKAAYGGVGENVLLYV--HADDFVEFYV   71
        |||::||::||::||::||::||::||::||::||::||::||::||::||::||::||::||::||
Db     1048 WEVFEEQALKECEFEVN-----SSVDRCIY-IIDKNIGNINENITVILLYYDDVIAT    1095
```

```

Qy      72 GESLKN-----VRNLDVMPIFETLALRLVLOGDVIWL 104
      | : | | : | : : : | | : |
Db      1096 GDMTRMNNFKRYLMEKFRMTDLNEIKHFIGIRIEMQEDKIYL 1137

```

## RESULT 14

```

probable siderophore biosyntheas protein ypo0778 [imported] - Yersinia pestis (strain
C:Species: Yersinia pestis
#Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 02-Nov-2001

```

R:Parkhill, J.Wren, B.W.; Thomson, N.R.; Ritball, R.W.; Holden, M.T.G.; Prentice, M.B.  
deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;  
Lil, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrrell,

A;Title: Genome sequence of *Yersinia pestis*, the causative agent of plague  
A;Reference number: AB0001; NCID:21470413; PMID:11566360  
A;Accession: AH0095

A;Molecule type: DNA  
A;Residues: 1-2201 <KUR>  
A;Cross-references: GB:AL590842; PIDN:CAC089627.1; PID:g15978856; GSPDB:GN00175

A;Gene: YPO0778  
Query Match: 11.3%; Score 71; DB 2; Length 2201;

Matches	33;	Conservative	24;	Mismatches	26;	Indels	34;	Gaps	8;
Qy	21	EDNRCYIDNGK--LPARGSTIVGNRSRF-VFDPKADYGVGVENLIYV-HADVVEFVFGESLK	76						

```

D6 911 E1NR1YVDGK1L1PAL1AL1FDLSVYD1FGPL1---VGG1SLV1PNAGDEK---EAKQ 962
QY 77 W-----NVRNLDVMP-1FETAL-----RLVEQGV1WLRCPBELR 111

```

Dd 963 WLSALHGHQVTHMNSVPALFEMLLIAEGCTQALPRSLQGVLLSGDWIGDILPRLR 1019

## phosphoma

C/SpecArch: AgriOdontocetium lunataeArchiv  
C/Date: 11-Jan-2002 #sequence\_revision 11-Jan-2002 #text\_change 18-Nov-2002  
C/Accession: A22963  
R/Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I.  
erage, G.; Gillet, P.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McCellen, S.  
Science 294, 2317-2323, 2001  
A/Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,  
ter, E.W.

A:Reference number: AB2577; MUID:21608550; PMID:117431933

A; ACCESSION: AF2963

A; status: preliminary

A; molecule type: DNA

A; Residues: 1-334 (NOK)

MULTI-TECHNICAL: SP:HEAVYWEIGHT, FIRM:MUTUALITY, FIN:GT/10000, SECUR:GAINSTO

A;Experimental source: strain C58 (Dupont  
C;Genetics:  
A;Gene: Atu3311  
A;Map position: linear chromosome

Query Match	11.2%;	Score 70.5;	DB 2;	Length 394;
Best Local Similarity	27.1%;	Pred. No. 14;		
Matches 23;	Conservative 12;	Mismatches 39;	Indels 11;	Gaps 5;

QY 8 PMKSTAKH--PMFOIEDNRCY-IDNGKLFAFGSIIGNMSKFVEDPRADYGIGENLYVHA 64  
||| ||| : |::||| ::| |::|  
Db 82 PMXDAISHGLSWFE---KYRLENG--LYGNLADQTGRII-DPSFDLLYNQAFAALFAAA 133

QY 65 DVEFVPGESLKNVRLDVMPIFE 89  
: | | | |  
Db 134 QAAASLPERKDEMRVRLDILATLE 158

Search completed: January 22, 2004, 11:42:39  
Job time : 23.3788 secs

Y\_28\_143.rpr

Page 5



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Comugen Ltd.

OM protein - protein search, using sw model

Run on: January 22, 2004, 11:41:40 ; Search time 431.324 Seconds  
(without alignments)  
84.837 Million cell updates/sec

Title: US-09-890-806-5  
Perfect score: 947  
Sequence: 1 MKFLVVALVFMVYISYIY.....GQPELAPEDPEDSALLEDEV 177

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 777136 seqs, 206736638 residues  
Total number of hits satisfying chosen parameters: 777136

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

- 1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*
- 2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*
- 8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*
- 10: /cgn2\_6/ptodata/2/pubpaa/US09B\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep.\*
- 12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*
- 13: /cgn2\_6/ptodata/2/pubpaa/US10A\_PUBCOMB.pep.\*
- 14: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pep.\*
- 15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep.\*
- 16: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*
- 17: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*
- 18: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	740	78.1	199	15	US-10-300-393-19 Sequence 19, Appl
2	117.5	12.4	451	14	US-10-029-009-20 Sequence 20, Appl
3	115	12.1	186	15	US-10-300-393-20 Sequence 20, Appl
4	110	11.6	488	14	US-10-029-009-22 Sequence 22, Appl
5	109.5	11.6	1189	15	US-10-044-692-325 Sequence 325, App
6	109.5	11.6	1189	15	US-10-044-539-325 Sequence 325, App
7	104	11.0	496	14	US-10-029-009-34 Sequence 34, Appl
8	102	10.8	21	9	US-09-766-378A-36 Sequence 36, Appl
9	102	10.8	21	11	US-09-927-122-3 Sequence 3, Appl
10	102	10.8	21	11	US-09-927-121B-3 Sequence 3, Appl
11	102	10.8	21	12	US-10-361-849-8 Sequence 8, Appl
12	102	10.8	21	14	US-10-029-009-24 Sequence 24, Appl
13	102	10.8	49	12	US-10-256-977-10 Sequence 10, Appl
14	102	10.8	49	15	US-10-084-298-10 Sequence 10, Appl
15	102	10.8	70	11	US-09-847-208-26 Sequence 26, Appl

16	102	10.8	129	15	US-10-152-190-6 Sequence 6, Appl
17	102	10.8	260	15	US-10-264-634-23 Sequence 23, Appl
18	102	10.8	260	15	US-10-264-634-39 Sequence 39, Appl
19	99	10.5	41	12	US-10-410-842A-40 Sequence 40, Appl
20	99	10.5	41	12	US-10-410-842A-44 Sequence 44, Appl
21	99	10.5	61	12	US-10-410-842A-30 Sequence 30, Appl
22	99	10.5	104	12	US-10-410-842A-6 Sequence 6, Appl
23	99	10.5	104	12	US-10-369-300-13 Sequence 13, Appl
24	99	10.5	394	12	US-10-410-842A-2 Sequence 2, Appl
25	97	10.2	20	14	US-10-029-009-12 Sequence 12, Appl
26	97	10.2	21	12	US-10-133-973-74 Sequence 74, Appl
27	91	9.6	41	12	US-10-410-842A-56 Sequence 56, Appl
28	91	9.6	41	12	US-10-410-842A-69 Sequence 69, Appl
29	86	9.1	6291	12	US-10-329-079-61 Sequence 41, Appl
30	82	8.7	893	14	US-10-014-436-4 Sequence 4, Appl
31	82	8.7	1938	14	US-10-014-436-2 Sequence 2, Appl
32	79	8.3	329	12	US-10-104-047-3709 Sequence 3709, Ap
33	79	8.3	3931	12	US-10-120-801-18 Sequence 18, Appl
34	76.5	8.1	316	10	US-09-736-457-340 Sequence 340, App
35	76.5	8.1	316	10	US-09-902-941-340 Sequence 340, App
36	76.5	8.1	316	10	US-09-849-626-340 Sequence 340, App
37	76.5	8.1	316	11	US-09-476-300-340 Sequence 340, App
38	76.5	8.1	316	12	US-10-274-375-3 Sequence 3, Appl
39	76.5	8.1	316	12	US-10-113-872-340 Sequence 340, App
40	76.5	8.1	316	15	US-10-017-754-340 Sequence 340, App
41	76.5	8.1	358	9	US-09-925-299-923 Sequence 923, App
42	76.5	8.1	358	11	US-09-925-299-923 Sequence 923, App
43	75.5	8.0	248	12	US-10-131-409-85 Sequence 85, Appl
44	75.5	8.0	248	12	US-10-150-811-85 Sequence 85, Appl
45	75.5	8.0	267	12	US-10-131-409-83 Sequence 83, Appl

ALIGNMENTS

RESULT 1  
US-10-300-393-19  
; Sequence 19, Application US/10300393  
; Publication No. US20030118568A1  
; GENERAL INFORMATION:  
; APPLICANT: Crew, Mark D.  
; TITLE OF INVENTION: Viral Stealth Technology to Prevent T-Cell-Mediated Rejection of  
; FILE REFERENCE: 8793-50391  
; CURRENT APPLICATION NUMBER: US/10/300,393  
; CURRENT FILING DATE: 2002-11-19  
; PRIOR APPLICATION NUMBER: US 60/342,981  
; PRIOR FILING DATE: 2001-12-18  
; NUMBER OF SEQ ID NOS: 22  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 19  
; LENGTH: 199  
; TYPE: PRT  
; ORGANISM: human cytomagalovirus Towne strain US2  
US-10-300-393-19

Query Match	Score	DB	Length	Best Local Similarity	Pred. No.	Matches	Conservative	Mismatches	Indels	Gaps
22	78.1%	100.0%	199	1.5e-77	0	136	0	0	0	0
20	78.1%	100.0%	199	1.5e-77	0	136	0	0	0	0
82	78.1%	100.0%	199	1.5e-77	0	136	0	0	0	0
80	78.1%	100.0%	199	1.5e-77	0	136	0	0	0	0
142	78.1%	100.0%	199	1.5e-77	0	136	0	0	0	0
140	78.1%	100.0%	199	1.5e-77	0	136	0	0	0	0



APPLICATION NUMBER: US 08/851,843  
 FILING DATE: 06-MAY-1997  
 APPLICATION NUMBER: US 08/846,017  
 FILING DATE: 25-APR-1997  
 APPLICATION NUMBER: US 08/844,419  
 FILING DATE: 18-APR-1997  
 APPLICATION NUMBER: US 08/724,643  
 FILING DATE: 01-OCT-1996  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Apple, Randolph T.  
 REGISTRATION NUMBER: 36,429  
 REFERENCE/DOCKET NUMBER: 015389-002600US  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (415) 576-0200  
 TELEFAX: (415) 576-0300  
 INFORMATION FOR SEQ ID NO: 325:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1189 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: <Unknown>  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 SEQUENCE DESCRIPTION: SEQ ID NO: 325:  
 US-10-044-692-325  
 Query Match 11.6%; Score 109.5; DB 15; Length 1189;  
 Best Local Similarity 32.1%; Pred. No. 0.0026;  
 Matches 50; Conservative 10; Mismatches 65; Indels 31; Gaps 6;  
 QY 1 MKELVNVALVFVWVYISYIYARLPDGIKAG--EDA-----LRPWKSTAKHPWFOI 49  
 DB 1 MKELVNVALVFVWVYISYIYAD-PSRSAAGTMEFAASTORCVLLRTWEALAPATPAMP 59  
 QY 50 EDNRCTYDNGKLFAGSTYGNMRFVFPDKADYGVGENLYVHADVER-----V 99  
 DB 60 RAPRCRAVRSLSRSHREVLPATFV---RRIGQGMRLVGRGPAFRALVAQCLVCV 115  
 QY 100 PGEISLKMVNRNDVMPFEFTLRLVLOGDVIWLRG 135  
 DB 116 P-----WDARPPPAAPSPFQVSCLELVARVLQRIC 146  
 RESULT 6  
 US-10-044-539-325  
 Sequence 325, Application US/10044539  
 Publication No. US2003010093A1  
 GENERAL INFORMATION:  
 APPLICANT: Cech, Thomas R.  
 Lingner, Joachim  
 Nakamura, Toru  
 Chapman, Karen B.  
 Morin, Gregg B.  
 Harley, Calvin  
 Andrews, William H.  
 TITLE OF INVENTION: HUMAN TELOMERASE CATALYTIC SUBUNIT: DIAGNOSTIC AND  
 THERAPEUTIC METHODS  
 NUMBER OF SEQUENCES: 335  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Townsend and Townsend and Crew LLP  
 STREET: Two Embarcadero Center, 8th Floor  
 CITY: San Francisco  
 STATE: California  
 COUNTRY: United States of America  
 ZIP: 94111  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patentin Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/10/044,539  
 FILING DATE: 11-Jan-2002  
 CLASSIFICATION: 435

PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 08/912,951  
 FILING DATE: <Unknown>  
 APPLICATION NUMBER: US 08/854,050  
 FILING DATE: 09-MAY-1997  
 APPLICATION NUMBER: US 08/851,843  
 FILING DATE: 06-MAY-1997  
 APPLICATION NUMBER: US 08/846,017  
 FILING DATE: 25-APR-1997  
 APPLICATION NUMBER: US 08/844,419  
 FILING DATE: 18-APR-1997  
 APPLICATION NUMBER: US 08/724,643  
 FILING DATE: 01-OCT-1996  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Apple, Randolph T.  
 REGISTRATION NUMBER: 36,429  
 REFERENCE/DOCKET NUMBER: 015389-002600US  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (415) 576-0200  
 TELEFAX: (415) 576-0300  
 INFORMATION FOR SEQ ID NO: 325:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1189 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: <Unknown>  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 SEQUENCE DESCRIPTION: SEQ ID NO: 325:  
 US-10-044-539-325  
 Query Match 11.6%; Score 109.5; DB 15; Length 1189;  
 Best Local Similarity 32.1%; Pred. No. 0.0026;  
 Matches 50; Conservative 10; Mismatches 65; Indels 31; Gaps 6;  
 QY 1 MKELVNVALVFVWVYISYIYARLPDGIKAG--EDA-----LRPWKSTAKHPWFOI 49  
 DB 1 MKELVNVALVFVWVYISYIYAD-PSRSAAGTMEFAASTORCVLLRTWEALAPATPAMP 59  
 QY 50 EDNRCTYDNGKLFAGSTYGNMRFVFPDKADYGVGENLYVHADVER-----V 99  
 DB 60 RAPRCRAVRSLSRSHREVLPATFV---RRIGQGMRLVGRGPAFRALVAQCLVCV 115  
 QY 100 PGEISLKMVNRNDVMPFEFTLRLVLOGDVIWLRG 135  
 DB 116 P-----WDARPPPAAPSPFQVSCLELVARVLQRIC 146  
 RESULT 7  
 US-10-029-009-34  
 Sequence 34, Application US/10029009  
 Publication No. US20020164617A1  
 GENERAL INFORMATION:  
 APPLICANT: Felsch, Jason S.  
 Applicant: Annis, David Allen  
 Applicant: Kalghatgi, Krishna  
 Applicant: Nash, Huw M.  
 TITLE OF INVENTION: Affinity Selection-Based Screening of Hydrophobic Proteins  
 FILE REFERENCE: 111025.173 US2  
 CURRENT APPLICATION NUMBER: US/10/029,009  
 CURRENT FILING DATE: 2002-03-28  
 PRIOR APPLICATION NUMBER: US 60/258,970  
 PRIOR FILING DATE: 2000-12-29  
 NUMBER OF SEQ ID NOS: 34  
 SOFTWARE: FastSeq for Windows Version 4.0  
 SEQ ID NO 34  
 LENGTH: 496  
 TYPE: PRT  
 ORGANISM: Artificial Sequence  
 FEATURE:  
 OTHER INFORMATION: Met-Melitin SS-FLAG M1 tag-m2 machr sequence  
 US-10-029-009-34  
 Query Match 11.0%; Score 104; DB 14; Length 496;



Best Local Similarity 100.0%; Pred. No. 8.4e-05;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKFLVVALVFMVVVYSIYA 21  
Db 1 MKFLVVALVFMVVVYSIYA 21

RESULT 12

US-10-029-009-24  
Sequence 24, Application US/10029009  
Publication No. US20020164617A1  
GENERAL INFORMATION:

APPLICANT: Felsch, Jason S.  
APPLICANT: Annis, David Allen  
APPLICANT: Kalphatgi, Krishna  
APPLICANT: Naei, Huw M.  
TITLE OF INVENTION: Affinity Selection-Based Screening of Hydrophobic Proteins  
FILE REFERENCE: 111025.173 US2  
CURRENT APPLICATION NUMBER: US/10/029,009  
CURRENT FILING DATE: 2002-03-28  
PRIOR APPLICATION NUMBER: US 60/258,970  
PRIOR FILING DATE: 2000-12-29  
NUMBER OF SEQ ID NOS: 34  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 24  
LENGTH: 21  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Melittin signal sequence  
US-10-029-009-24

Query Match 10.8%; Score 102; DB 14; Length 21;  
Best Local Similarity 100.0%; Pred. No. 8.4e-05;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKFLVVALVFMVVVYSIYA 21  
Db 1 MKFLVVALVFMVVVYSIYA 21

RESULT 13

US-10-256-977-10  
Sequence 10, Application US/10256977  
Publication No. US20030157106A1  
GENERAL INFORMATION:

APPLICANT: Jacobs, Kenneth  
APPLICANT: Pittman, Debra  
APPLICANT: Fouser, Lynette  
APPLICANT: Spaulding, Vikki  
APPLICANT: Xuan, Dejun  
TITLE OF INVENTION: Composition and Method for Treating Inflammatory  
FILE REFERENCE: G15358 CIP  
CURRENT APPLICATION NUMBER: US/10/256,977  
CURRENT FILING DATE: 2002-09-27  
PRIOR APPLICATION NUMBER: US/10/084,298  
PRIOR FILING DATE: 2002-09-10  
PRIOR APPLICATION NUMBER: 60/270,823  
PRIOR FILING DATE: 2001-02-23  
PRIOR APPLICATION NUMBER: 60/281,353  
PRIOR FILING DATE: 2001-04-03  
PRIOR APPLICATION NUMBER: 60/131,473  
PRIOR FILING DATE: 1999-04-28  
PRIOR APPLICATION NUMBER: 09/561,811  
PRIOR FILING DATE: 2000-04-28  
NUMBER OF SEQ ID NOS: 10  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 10  
LENGTH: 49  
TYPE: PRT  
ORGANISM: Artificial Sequence

FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence: Amino acid tag  
US-10-256-977-10

Query Match 10.8%; Score 102; DB 12; Length 49;  
Best Local Similarity 100.0%; Pred. No. 0.00026;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKFLVVALVFMVVVYSIYA 21  
Db 1 MKFLVVALVFMVVVYSIYA 21

RESULT 14

US-10-084-298-10  
Sequence 10, Application US/10084298  
Publication No. US2003009649A1  
GENERAL INFORMATION:

APPLICANT: Jacobs, Kenneth  
APPLICANT: Pittman, Debra  
APPLICANT: Fouser, Lynette  
APPLICANT: Spaulding, Vikki  
APPLICANT: Xuan, Dejun  
TITLE OF INVENTION: Composition and Method for Treating Inflammatory  
FILE REFERENCE: G15358 CIP  
CURRENT APPLICATION NUMBER: US/10/084,298  
CURRENT FILING DATE: 2002-09-10  
PRIOR APPLICATION NUMBER: 60/270,823  
PRIOR FILING DATE: 2001-02-23  
PRIOR APPLICATION NUMBER: 60/281,353  
PRIOR FILING DATE: 2001-04-03  
PRIOR APPLICATION NUMBER: 60/131,473  
PRIOR FILING DATE: 1999-04-28  
PRIOR APPLICATION NUMBER: 09/561,811  
PRIOR FILING DATE: 2000-04-28  
NUMBER OF SEQ ID NOS: 10  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 10  
LENGTH: 49  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence: Amino acid tag  
US-10-084-298-10

Query Match 10.8%; Score 102; DB 15; Length 49;  
Best Local Similarity 100.0%; Pred. No. 0.00026;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKFLVVALVFMVVVYSIYA 21  
Db 1 MKFLVVALVFMVVVYSIYA 21

RESULT 15

US-09-847-208-26  
Sequence 26, Application US/09847208  
Publication No. US20030082190A1  
GENERAL INFORMATION:

APPLICANT: Saxon, Andrew  
APPLICANT: Zhang, Ke  
APPLICANT: Zhu, Daocheng  
TITLE OF INVENTION: FUSION MOLECULES AND TREATMENT OF  
TITLE OF INVENTION: 19E-MEDIATED ALLERGIC DISEASES  
FILE REFERENCE: UC67,002A  
CURRENT APPLICATION NUMBER: US/09/847,208  
CURRENT FILING DATE: 2001-05-01  
NUMBER OF SEQ ID NOS: 177  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 26  
LENGTH: 70  
TYPE: PRT  
ORGANISM: Artificial Sequence

! ORGANISM: Apis mellifera(Honeybee)Apis cerana(Ind. honeybee)  
US-09-847-208-26

Query Match 10.8%; Score 102; DB 11; Length 70;  
Best Local Similarity 100.0%; Pred. No. 0.00043;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKFLNNVALVPMVVYISYIA 21  
Db 1 MKFLNNVALVPMVVYISYIA 21

Search completed: January 22, 2004, 12:09:28  
Job time : 431.324 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 22, 2004, 11:37:25 ; Search time 33.2253 Seconds  
(without alignments)  
225.401 Million cell updates/sec

Title: US-09-890-806-5  
Perfect score: 947  
Sequence: 1 MKFLVVALVFMVYISYIY.....GQPLAPDEPDSDALLEDPV 177

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA:\*  
1: /cgn2\_6/prodata/2/1aa/5A\_COMB.pep:\*  
2: /cgn2\_6/prodata/2/1aa/5B\_COMB.pep:\*  
3: /cgn2\_6/prodata/2/1aa/5A\_COMB.pep:\*  
4: /cgn2\_6/prodata/2/1aa/5B\_COMB.pep:\*  
5: /cgn2\_6/prodata/2/1aa/PCTUS\_COMB.pep:\*  
6: /cgn2\_6/prodata/2/1aa/Backfillseq1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	109.5	11.6	1189	3	US-08-974-549A-613 Sequence 613, App
2	109.5	11.6	1189	4	US-08-912-951-325 Sequence 325, App
3	102	10.8	21	3	US-08-960-190A-36 Sequence 36, App1
4	102	10.8	21	4	US-09-376-330-4 Sequence 4, App1
5	99	10.5	308	1	US-08-499-568-2 Sequence 2, App1
6	99	10.5	308	2	US-08-793-958-2 Sequence 2, App1
7	99	10.5	369	2	US-08-139-609-1 Sequence 1, App1
8	99	10.5	383	6	5470718-5 Patent No. 5470718
9	99	10.5	388	1	US-08-499-568-11 Sequence 11, App1
10	99	10.5	388	2	US-08-793-958-11 Sequence 11, App1
11	99	10.5	383	1	US-08-220-151-23 Sequence 23, App1
12	99	10.5	383	1	US-08-413-118-23 Sequence 23, App1
13	99	10.5	393	3	US-08-473-446-23 Sequence 23, App1
14	99	10.5	393	6	5182195-10 Patent No. 5182195
15	99	10.5	394	1	US-08-499-568-4 Sequence 4, App1
16	99	10.5	394	2	US-08-793-958-4 Sequence 4, App1
17	97	10.2	21	1	US-08-334-669-2 Sequence 2, App1
18	97	10.2	21	2	US-08-485-341A-2 Sequence 2, App1
19	91	9.6	368	5	PCT-US93-11703-24 Sequence 24, App1
20	91	9.6	368	1	US-08-499-568-15 Sequence 15, App1
21	91	9.6	393	2	US-08-793-958-15 Sequence 15, App1
22	91	9.6	393	2	US-08-956-998-2 Sequence 2, App1
23	87.5	9.2	26	3	US-08-782-480-43 Sequence 43, App1
24	87.5	9.2	26	3	US-08-954-211-43 Sequence 43, App1
25	87.5	9.2	26	4	US-09-005-167A-43 Sequence 43, App1
26	87.5	9.2	26	4	US-09-176-741B-43 Sequence 43, App1
27	82	8.7	893	4	US-09-514-302-4 Sequence 4, App1

28	82	8.7	1938	4	US-09-514-302-2 Sequence 2, App1
29	80	8.4	676	3	US-08-947-965-71 Sequence 71, App1
30	78	8.2	310	4	US-09-252-991A-17381 Sequence 17381, A
31	77.5	8.2	729	4	US-09-252-991A-32535 Sequence 32535, A
32	76.5	8.1	316	4	US-09-702-705-340 Sequence 340, App
33	76.5	8.1	316	4	US-09-736-457-340 Sequence 340, App
34	72	7.6	675	3	US-08-947-965-76 Sequence 76, App1
35	71.5	7.6	1266	4	US-08-506-296B-4 Sequence 28582, A
36	71	7.5	405	4	US-09-252-991A-28582 Sequence 28582, A
37	70.5	7.4	327	4	US-09-107-532A-6181 Sequence 6181, AP
38	70	7.4	448	4	US-09-724-623-84 Sequence 84, App1
39	69.5	7.3	358	1	US-08-604-913B-11 Sequence 11, App1
40	69.5	7.3	472	4	US-08-985-492-13 Sequence 13, App1
41	69.5	7.3	521	1	US-08-276-213-3 Sequence 3, App1
42	68.5	7.2	130	4	US-08-826-134-12 Sequence 12, App1
43	68.5	7.2	640	4	US-09-252-991A-27542 Sequence 27542, A
44	68.5	7.2	1384	4	US-08-826-134-2 Sequence 2, App1
45	68	7.2	316	1	US-08-585-595-3 Sequence 3, App1

# ALIGNMENTS

RESULT 1  
US-08-974-549A-613  
; Sequence 613, Application US/08974549A  
; Patent No. 6166178  
; GENERAL INFORMATION:  
; APPLICANT: Cech, Thomas R.  
; APPLICANT: Langer, Joachim  
; APPLICANT: Nakamura, Toru  
; APPLICANT: Chapman, Karen B.  
; APPLICANT: Morin, Gregg B.  
; APPLICANT: Harley, Calvin B.  
; APPLICANT: Andrews, William H.  
; TITLE OF INVENTION: Human Telomerase Catalytic Subunit  
; NUMBER OF SEQUENCES: 727  
; CORRESPONDENCE ADDRESS:  
; ADDRESSER: Townsend and Townsend and Crew LLP  
; STREET: Two Embarcadero Center, Eighth Floor  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94111-3834  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/974,549A  
; FILING DATE: 19-NOV-1997  
; CLASSIFICATION: 536  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/724,643  
; FILING DATE: 01-OCT-1996  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/844,419  
; FILING DATE: 18-APR-1997  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/846,017  
; FILING DATE: 25-APR-1997  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/851,843  
; FILING DATE: 06-MAY-1997  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/854,050  
; FILING DATE: 09-MAY-1997  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/911,312  
; FILING DATE: 14-AUG-1997  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/912,951



CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Dikey, Bronstein, Roberts & Cushman, LLP  
 STREET: 130 Water Street  
 CITY: Boston  
 STATE: MA  
 COUNTRY: usa  
 ZIP: 02109  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Diskette  
 OPERATING SYSTEM: IBM Compatible  
 SOFTWARE: FASTSEQ for Windows Version 2.0  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/960,190A  
 FILING DATE: 29-OCT-1997  
 CLASSIFICATION: 536  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER:  
 FILING DATE:  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Corless, Peter F.  
 REGISTRATION NUMBER: 33,860  
 REFERENCE/DOCKET NUMBER: 48002  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 617-523-3400  
 TELEFAX: 617-523-6440  
 TELEX:  
 INFORMATION FOR SEQ ID NO: 36:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 21 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-08-960-190A-36  
 Query Match 10.8%; Score 102; DB 3; Length 21;  
 Best local Similarity 100.0%; Pred. No. 1.2e-05;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MKFLVNVALVFVWVYISYIYA 21  
 Db 1 MKFLVNVALVFVWVYISYIYA 21  
 RESULT 4  
 US-09-376-330-4  
 Sequence 4, Application US/09376330  
 Patent No. 639321  
 GENERAL INFORMATION:  
 APPLICANT: Tessier, Daniel C.  
 APPLICANT: Dignard, Daniel  
 APPLICANT: Bergeron, John J.M.  
 TITLE OF INVENTION: Method for screening for  
 TITLE OF INVENTION: UDP-glucose:glycoprotein glucosyltransferase (UGGT) activity  
 TITLE OF INVENTION: and nucleic acid encoding for UGGT  
 FILE REFERENCE: 2139-9"US"  
 CURRENT APPLICATION NUMBER: US/09/376,330  
 CURRENT FILING DATE: 1999-08-18  
 NUMBER OF SEQ ID NOS: 35  
 SOFTWARE: FASTSEQ for Windows Version 3.0  
 SEQ ID NO 4  
 LENGTH: 21  
 TYPE: PRT  
 ORGANISM: Melittin signal peptide sequence  
 US-09-376-330-4  
 Query Match 10.8%; Score 102; DB 4; Length 21;  
 Best local Similarity 100.0%; Pred. No. 1.2e-05;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MKFLVNVALVFVWVYISYIYA 21

Db 1 MKFLVNVALVFVWVYISYIYA 21  
 RESULT 5  
 US-08-499-568-2  
 Sequence 2, Application US/08499568  
 Patent No. 5654174  
 GENERAL INFORMATION:  
 APPLICANT: Cohen, Gary H.  
 APPLICANT: Eisenberg, Roselyn J.  
 APPLICANT: Nicola, Anthony  
 TITLE OF INVENTION: Herpes Simplex Virus Glycoprotein D  
 TITLE OF INVENTION: Variants  
 NUMBER OF SEQUENCES: 15  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
 STREET: 6300 Sears Tower, 233 S. Wacker Drive  
 CITY: Chicago  
 STATE: Illinois  
 COUNTRY: USA  
 ZIP: 60606  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 OPERATING SYSTEM: IBM PC Compatible  
 SOFTWARE: Patent Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/499,568  
 FILING DATE:  
 CLASSIFICATION: 424  
 ATTORNEY/AGENT INFORMATION:  
 NAME: No. 5654174and, Greta E.  
 REGISTRATION NUMBER: 35,302  
 REFERENCE/DOCKET NUMBER: 32813  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (312) 474-6300  
 TELEFAX: (312) 474-0448  
 TELEX: 25-3856  
 INFORMATION FOR SEQ ID NO: 2:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 308 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-08-499-568-2  
 Query Match 10.5%; Score 99; DB 1; Length 308;  
 Best local Similarity 100.0%; Pred. No. 0.0015;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 159 QPELAPDEPDSDALIEDPV 177  
 Db 267 QPELAPDEPDSDALIEDPV 285  
 RESULT 6  
 US-08-793-958-2  
 Sequence 2, Application US/08793958  
 Patent No. 581486  
 GENERAL INFORMATION:  
 APPLICANT: Cohen, Gary H.  
 APPLICANT: Eisenberg, Roselyn J.  
 APPLICANT: Nicola, Anthony  
 TITLE OF INVENTION: Herpes Simplex Virus Glycoprotein D  
 TITLE OF INVENTION: Variants  
 NUMBER OF SEQUENCES: 15  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
 STREET: 6300 Sears Tower, 233 S. Wacker Drive  
 CITY: Chicago  
 STATE: Illinois  
 COUNTRY: USA

ZIP: 60606  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/793,958  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/499,568  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: No.581446and, Greta E.  
REGISTRATION NUMBER: 35,302  
REFERENCE/DOCKET NUMBER: 32813  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (312) 474-6300  
TELEFAX: (312) 474-0448  
TELEX: 25-3856  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 308 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-793-958-2

Query Match 10.5%; Score 99; DB 2; Length 308;  
Best Local Similarity 100.0%; Pred. No. 0.0015;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 159 QPELAPDEPDSALLDPV 177  
Db 267 QPELAPDEPDSALLDPV 285

RESULT 7  
US-08-139-609-1  
Sequence 1, Application US/08139609  
Patent No. 5837249  
GENERAL INFORMATION:  
APPLICANT: Heber-Katz, Ellen  
APPLICANT: Dietzschold, Bernhard  
TITLE OF INVENTION: Method for Generating an Immunogenic T  
TITLE OF INVENTION: Cell Response Protective Against a Virus  
NUMBER OF SEQUENCES: 13  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Howson and Howson  
STREET: Spring House Corporate Cntr., P.O. Box 457  
CITY: Spring House  
STATE: Pennsylvania  
COUNTRY: USA  
ZIP: 19477  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/139,609  
FILING DATE:  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/868,946  
FILING DATE: 15-APR-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/685,459  
FILING DATE: 12-APR-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/047,443  
FILING DATE: 08-MAY-1987

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 06/725,087  
FILING DATE: 19-APR-1985  
ATTORNEY/AGENT INFORMATION:  
NAME: Bak, Mary E.  
REGISTRATION NUMBER: 31,215  
REFERENCE/DOCKET NUMBER: WST11DUSA  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 215-540-9200  
TELEFAX: 215-540-5818  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 369 amino acids  
TYPE: amino acid  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide  
US-08-139-609-1

Query Match 10.5%; Score 99; DB 2; Length 369;  
Best Local Similarity 100.0%; Pred. No. 0.002;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 159 QPELAPDEPDSALLDPV 177  
Db 265 QPELAPDEPDSALLDPV 283

RESULT 8  
5470718-5  
Patent No. 5470718  
APPLICANT: O'CALLAGHAN, DENNIS J.  
TITLE OF INVENTION: EQUINE HERPESVIRUS TYPE 1 GLYCOPROTEIN  
D NUCLEIC ACIDS  
NUMBER OF SEQUENCES: 8  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/954,417  
FILING DATE: 30-SEP-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 561,553  
FILING DATE: 01-AUG-1990  
SEQ ID NO:5  
LENGTH: 383  
5470718-5

Query Match 10.5%; Score 99; DB 6; Length 383;  
Best Local Similarity 100.0%; Pred. No. 0.0021;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 159 QPELAPDEPDSALLDPV 177  
Db 291 QPELAPDEPDSALLDPV 309

RESULT 9  
US-08-499-568-11  
Sequence 11, Application US/08499568  
Patent No. 5654174  
GENERAL INFORMATION:  
APPLICANT: Conen, Gary H.  
APPLICANT: Eisenberg, Roselyn J.  
APPLICANT: Nicola, Anthony  
TITLE OF INVENTION: Herpes Simplex Virus Glycoprotein D  
TITLE OF INVENTION: Variants  
NUMBER OF SEQUENCES: 15  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
STREET: 6300 Sears Tower, 233 S. Wacker Drive  
CITY: Chicago  
STATE: Illinois  
COUNTRY: USA  
ZIP: 60606  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/499,568  
FILING DATE:  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: No. 5654174and, Greta E.  
REGISTRATION NUMBER: 35,302  
REFERENCE/DOCKET NUMBER: 32813  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (312) 474-6300  
TELEFAX: (312) 474-0448  
TELEX: 25-3856  
SEQUENCE CHARACTERISTICS:  
LENGTH: 388 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-499-568-11

Query Match 10.5%; Score 99; DB 1; Length 388;  
Best Local Similarity 100.0%; Pred. No. 0.0022;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 159 OPELAPEDPEDSALLEDPV 177  
DB 290 OPELAPEDPEDSALLEDPV 308

RESULT 10  
US-08-793-958-11  
Sequence 11, Application US/08793958  
Patent No. 5814486  
GENERAL INFORMATION:  
APPLICANT: Cohen, Gary H.  
APPLICANT: Eisenberg, Roselyn J.  
APPLICANT: Nicola, Anthony  
TITLE OF INVENTION: Herpes Simplex Virus Glycoprotein D  
TITLE OF INVENTION: Variants  
NUMBER OF SEQUENCES: 15  
CORRESPONDENCE ADDRESS:  
ADDRESSER: Marshall, O'Toole, Gerstein, Murray & Borun  
STREET: 6300 Sears Tower, 233 S. Wacker Drive  
CITY: Chicago  
STATE: Illinois  
COUNTRY: USA  
ZIP: 60606  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/793,958  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/499,568  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: No. 5814486and, Greta E.  
REGISTRATION NUMBER: 35,302  
REFERENCE/DOCKET NUMBER: 32813  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (312) 474-6300  
TELEFAX: (312) 474-0448  
TELEX: 25-3856  
INFORMATION FOR SEQ ID NO: 11:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 388 amino acids

TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-793-958-11

Query Match 10.5%; Score 99; DB 2; Length 388;  
Best Local Similarity 100.0%; Pred. No. 0.0022;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 159 OPELAPEDPEDSALLEDPV 177  
DB 290 OPELAPEDPEDSALLEDPV 308

RESULT 11  
US-08-220-151-23  
Sequence 23, Application US/08220151  
Patent No. 5529780  
GENERAL INFORMATION:  
APPLICANT: Paolletti, Enzo  
APPLICANT: Lambach, Keith J.  
TITLE OF INVENTION: NUCLEOTIDE AND AMINO ACID SEQUENCES OF  
CANINE HERPESVIRUS 9B, 9C AND 9D AND USES THEREFOR  
NUMBER OF SEQUENCES: 91  
CORRESPONDENCE ADDRESS:  
ADDRESSER: Curtis, Morris & Safford  
STREET: 530 Fifth Avenue  
CITY: New York  
STATE: NY  
COUNTRY: USA  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/220,151  
FILING DATE: 30-MAR-1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Frommet, William S.  
REGISTRATION NUMBER: 25,506  
REFERENCE/DOCKET NUMBER: 454310-2540  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 840-3333  
TELEFAX: (212) 840-0712  
TELEX: 425066 CURTMS  
INFORMATION FOR SEQ ID NO: 23:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 393 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FRAGMENT TYPE: N-terminal  
US-08-220-151-23

Query Match 10.5%; Score 99; DB 1; Length 393;  
Best Local Similarity 100.0%; Pred. No. 0.0022;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 159 OPELAPEDPEDSALLEDPV 177  
DB 289 OPELAPEDPEDSALLEDPV 307

RESULT 12  
US-08-413-118-23  
Sequence 23, Application US/08413118  
Patent No. 5688920  
GENERAL INFORMATION:  
APPLICANT: PAOLETTI, ENZO

APPLICANT: LIMBACH, KEITH J.  
TITLE OF INVENTION: NUCLEOTIDE AND AMINO ACID SEQUENCES OF  
TITLE OF INVENTION: CANINE HERPESVIRUS gB, gC, AND gD AND USES THEREFOR  
NUMBER OF SEQUENCES: 128  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: CURTIS, MORRIS & SAFFORD, P.C.  
STREET: 530 FIFTH AVENUE, 25TH FLOOR  
CITY: NEW YORK  
STATE: NEW YORK  
COUNTRY: UNITED STATES OF AMERICA  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/413,118  
FILING DATE: 29-MAR-1995  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/220,151  
FILING DATE: 30-MAR-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: FROMMER, WILLIAM S.  
REGISTRATION NUMBER: 25,506  
REFERENCE/DOCKET NUMBER: 454310-2670  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 840-3333  
TELEFAX: (212) 840-0712  
INFORMATION FOR SEQ ID NO: 23:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 393 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FRAGMENT TYPE: N-terminal  
US-08-413-118-23

Query Match 10.5%; Score 99; DB 1; Length 393;  
Best Local Similarity 100.0%; Pred. No. 0.0022;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 159 QPELAPEDPDSALIEDPV 177  
Db 289 QPELAPEDPDSALIEDPV 307

RESULT 13  
US-08-473-446-23  
Sequence 23, Application US/08473446  
Patent No. 6017542  
GENERAL INFORMATION:  
APPLICANT: PAOLETTI, ENZO  
APPLICANT: LIMBACH, KEITH J.  
TITLE OF INVENTION: NUCLEOTIDE AND AMINO ACID SEQUENCES OF  
TITLE OF INVENTION: CANINE HERPESVIRUS gB, gC, AND gD AND USES THEREFOR  
NUMBER OF SEQUENCES: 128  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: CURTIS, MORRIS & SAFFORD, P.C.  
STREET: 530 FIFTH AVENUE, 25TH FLOOR  
CITY: NEW YORK  
STATE: NEW YORK  
COUNTRY: UNITED STATES OF AMERICA  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/473,446

FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/413,118  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: FROMMER, WILLIAM S.  
REGISTRATION NUMBER: 25,506  
REFERENCE/DOCKET NUMBER: 454310-2670  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 840-3333  
TELEFAX: (212) 840-0712  
INFORMATION FOR SEQ ID NO: 23:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 393 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FRAGMENT TYPE: N-terminal  
US-08-473-446-23

Query Match 10.5%; Score 99; DB 3; Length 393;  
Best Local Similarity 100.0%; Pred. No. 0.0022;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 159 QPELAPEDPDSALIEDPV 177  
Db 289 QPELAPEDPDSALIEDPV 307

RESULT 14  
5182195-10  
Patent No. 5182195  
APPLICANT: NAKAHARA, KAZUO;KAISHO, YOSHITAKO;YOSHIMURA, KOJI  
TITLE OF INVENTION: METHOD FOR INCREASING USING PROTEASE  
DEFICIENT YEASTS  
NUMBER OF SEQUENCES: 71  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/269,140  
FILING DATE: 09-NOV-1988  
SEQ ID NO:10:  
LENGTH: 393  
5182195-10

Query Match 10.5%; Score 99; DB 6; Length 393;  
Best Local Similarity 100.0%; Pred. No. 0.0022;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 159 QPELAPEDPDSALIEDPV 177  
Db 289 QPELAPEDPDSALIEDPV 307

RESULT 15  
US-08-499-568-4  
Sequence 4, Application US/08499568  
Patent No. 5654174  
GENERAL INFORMATION:  
APPLICANT: Cohen, Gary H.  
APPLICANT: Eisenberg, Roselyn J.  
APPLICANT: Nicola, Anthony  
TITLE OF INVENTION: Herpes Simplex Virus Glycoprotein D  
TITLE OF INVENTION: Variants  
NUMBER OF SEQUENCES: 15  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
STREET: 6300 Sears Tower, 233 S. Wacker Drive  
CITY: Chicago  
STATE: Illinois  
COUNTRY: USA  
ZIP: 60606  
COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/499,568  
 FILING DATE:  
 CLASSIFICATION: 424  
 ATTORNEY/AGENT INFORMATION:  
 NAME: No. 5654174and, Greta E.  
 REGISTRATION NUMBER: 35,302  
 REFERENCE/DOCKET NUMBER: 32813  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (312) 474-6300  
 TELEFAX: (312) 474-0448  
 TELEX: 25-3856  
 INFORMATION FOR SEQ. ID NO: 4:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 394 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-08-499-568-4

Query Match 10.5%; Score 99; DB 1; Length 394;  
 Best local Similarity 100.0%; Pred. No. 0.0022;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 159 OPELAPDEPDSALIEDPV 177  
 |||||  
 DB 290 OPELAPDEPDSALIEDPV 308

Search completed: January 22, 2004, 11:38:43  
 Job time : 34.2253 secs





GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: January 22, 2004, 11:37:25 ; Search time 32.6212 Seconds  
(without alignments)  
521.804 Million cell updates/sec

Title: US-09-890-806-5

Sequence: 1 MKFLVVALVFMVYISYIY.....GQELAPDEPDSALLEDPV 177

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: p1r1:\*  
2: p1r2:\*  
3: p1r3:\*  
4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	740	78.1	199	1 QOBECS	HQLF2 protein prec
2	125	13.2	186	1 QOBECS	HQLF2 precursor -
3	102	10.8	70	1 MPRH1	melittin, major, p
4	99	10.5	393	1 VGBBDZ	glycoprotein D pre
5	99	10.5	394	1 VGBBD1	glycoprotein D - h
6	99	10.5	394	1 VGBB17	glycoprotein D pre
7	96.5	10.2	394	1 A47627	glycoprotein D pre
8	96.5	10.2	1409	1 OFFPCP	glycoprotein D pre
9	91	9.6	393	1 VGBBD2	glycoprotein D - h
10	91	9.6	393	1 VGBBD3	glycoprotein D - h
11	91	9.6	393	2 E43674	US6 protein - huma
12	84.5	8.9	787	2 PC1232	copla polyprotein
13	84.5	8.9	1578	2 AD1512	peptidoglycan bou
14	84	8.9	159	2 F75021	ribosomal protein
15	84	8.9	254	2 AD0205	probable N-acetyl
16	81	8.6	410	2 S28354	hypothetical prote
17	80.5	8.5	415	2 T21532	hypothetical prote
18	80.5	8.5	1582	2 AC1153	adhesin homolog 1m
19	80	8.4	703	1 ALBSX1	cyclomaltodextrin
20	79	8.3	159	1 E71208	hypothetical prote
21	79	8.3	303	2 AI3115	conserved hypothet
22	78.5	8.3	404	2 S56335	hypothetical 43.7K
23	78.5	8.3	404	2 E91287	hypothetical prote
24	78.5	8.3	404	2 H86128	hypothetical prote
25	77.5	8.2	660	2 S71276	beta-fructofuranos
26	77.5	8.2	664	2 B86257	beta-fructosidase
27	77	8.1	272	2 T44528	phosphoglycolate p
28	76.5	8.1	248	1 A46542	granulin - Trichop
29	76	8.0	407	2 C82804	conserved hypothet

30	75.5	8.0	301	2 E72384	ABC transporter, A
31	75.5	8.0	316	2 A59021	aldehyde reductase
32	75.5	8.0	396	2 T04224	hypothetical prote
33	75.5	8.0	527	2 C70397	periplasmic cell d
34	75	7.9	394	2 AB1870	hypothetical prote
35	74.5	7.9	359	2 AI0140	molybdenum transpo
36	74	7.8	261	2 A98950	hypothetical prote
37	74	7.8	261	2 B85798	hypothetical prote
38	74	7.8	261	2 C64948	probable membrane
39	74	7.8	351	2 S11237	polymerase - Berne
40	74	7.8	360	2 A12854	conserved hypothet
41	74	7.8	360	2 F97631	hypothetical prote
42	74	7.8	473	1 T26280	linoleoyl-CoA deasa
43	73.5	7.8	557	1 T07909	4-coumarate-CoA li
44	73.5	7.8	757	2 B90572	lipoprotein limpor
45	73	7.7	463	2 D41853	hexose phosphate t

ALIGNMENTS

RESULT 1

QOBECS HQLF2 protein precursor - human cytomegalovirus (strain AD169)

N/Alternate names: hypothetical protein US2

C/Species: human cytomegalovirus, human herpesvirus 5

C/Date: 30-Sep-1989 #sequence\_revision 30-Sep-1989 #text\_change 28-Jul-2000

C/Accession: E26078; S09916

R/Weston, K.; Barrett, B.G.

J. Mol. Biol. 192, 177-208, 1986

A/Title: Sequence of the short unique region, short repeats, and part of the long repeats

A/Reference number: A92935; MUID:87169717; PMID:3031311

A/Accession: E26078

A/Molecule type: DNA

A/Residues: 1-199 <WES>

A/Cross-references: EMBL:X04650; NID:G59801; PIDN:CAB37096.1; PID:G4456177

A/Experimental source: strain AD169

R/Chen, M.S.; Bankier, A.T.; Beck, S.; Bohm, R.; Brown, C.M.; Cerny, R.; Hornselli, T.;

M.; Barrett, B.G.

Curr. Top. Microbiol. Immunol. 154, 125-169, 1990

A/Title: Analysis of the protein-coding content of the sequence of human cytomegalovirus

A/Reference number: S09749; MUID:90269039; PMID:2161319

A/Accession: S09916

A/Status: nucleic acid sequence not shown; translation not shown

A/Molecule type: DNA

A/Residues: 1-199 <CHR>

A/Cross-references: EMBL:X17403; NID:G59591; PIDN:CAA3513.1; PID:G1780933

A/Experimental source: strain AD169

A/Note: this sequence was submitted to the EMBL Data Library, December 1989

C/Genetics:

A/Gene: HQLF2

C/Superfamily: cytomegalovirus HQLF2 protein

C/Keywords: glycoprotein; transmembrane protein

F/1-25/Domain: signal sequence #status predicted <SIG>

F/26-199/Product: hypothetical protein US2 #status predicted <MAT>

F/161-186/Domain: transmembrane #status predicted <TM>

F/68/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match

Best Local Similarity 78.1%; Score 740; DB 1; Length 199;

Matches 136; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	22	RPPDITRAGEDALPWSKSTAGHPFQIEDNRCYIDNKLFRGSIYGMMSFVDPKAD	81
DB	20	RPPDITRAGEDALPWSKSTAGHPFQIEDNRCYIDNKLFRGSIYGMMSFVDPKAD	79
QY	82	YGVGENLYVHADVDFVPGESLKNVNRNLDVMP1FETLALRLVQGVYWLRCVPELRV	141
DB	80	YGVGENLYVHADVDFVPGESLKNVNRNLDVMP1FETLALRLVQGVYWLRCVPELRV	139
QY	142	DYTSAYMMNMQYGMV	157
DB	140	DYTSAYMMNMQYGMV	155

## RESULT 2

Q08EC6

HOLFI precursor - human cytomegalovirus (strain AD169)

N/Alternate names: hypothetical protein US3

C/Species: human cytomegalovirus, human herpesvirus 5

C/Date: 30-Sep-1989 #sequence\_revision 30-Sep-1989 #text\_change 28-Jul-2000

C/Accession: F26078; S09917

R/Weinst, K.; Barrell, B.G.

J. Mol. Biol. 192, 177-208, 1986

A/Title: Sequence of the short unique region, short repeats, and part of the long repeat

A/Reference number: A92935; MUID:87169717; PMID:3031311

A/Accession: F26078

A/Molecule type: DNA

A/Residues: 1-186 &lt;MES&gt;

A/Cross-references: EMBL:X04650; NID:959601; PIDN:CA837097.1; PID:94456178

A/Experimental source: strain AD169

R/Chen, M.S.; Bankier, A.T.; Beck, S.; Bohm, R.; Brown, C.M.; Cerny, R.; Horneill, T.;

M.; Barrell, B.G.

Curr. Top. Microbiol. Immunol. 154, 125-169, 1990

A/Title: Analysis of the protein-coding content of the sequence of human cytomegalovirus

A/Reference number: S09749; MUID:90269039; PMID:2161319

A/Accession: S09917

A/Status: nucleic acid sequence not shown; translation not shown

A/Molecule type: DNA

A/Residues: 1-186 &lt;CHE&gt;

A/Cross-references: EMBL:X17403; NID:959591; PIDN:CA835314.1; PID:91780934

A/Experimental source: strain AD169

A/Note: this sequence was submitted to the EMBL Data Library, December 1989

C/Genetics:

A/Gene: HOLFI

C/Superfamily: cytomegalovirus HOLFI protein

C/Keywords: glycoprotein, transmembrane protein

F/1-25/Domain: signal sequence #status predicted &lt;SIG&gt;

F/26-186/Product: hypothetical protein US3 #status predicted &lt;MAT&gt;

F/160-182/Domain: transmembrane #status predicted &lt;TMM&gt;

F/66/Binding site: carbohydrate (asn) (covalent) #status predicted

Query Match 13.24; Score 125; DB 1; Length 186;

Best Local Similarity 23.9%; Pred. No. 8.9e-05;

Matches 39; Conservative 36; Mismatches 56; Indels 32; Gaps 8;

QY 10 VFMVVISYIYARLPDGTITKAGEDALPFWKSTAGHPFOEDNRCYIDNGLFPARGISVG 69

Db 4 VLVALIAVLFLRLADSVPRPLDVV---SEISAHFRVENCCWFMGMLFKGKMSG 59

QY 70 NMSRFVDPKADYGVGSENL-----YVHADVFEVPGESLKNV-----RNLDM 114

Db 60 NFEKHF---VNGIVGQSVMDRLQVSGEQYHDE---RGAYFEMNIGGHPVTHVDV 112

QY 115 PIFETLALRLVQGD-VIWRFCVPELRVDYSSAYMMNMQYGM 156

Db 113 DI--TLSTR---WQDPKYACVQVMDYSSQTINWYLQPSM 150

## RESULT 3

MPH11 melittin, major precursor [validated] - honeybee

N/Comments: melittin F

C/Species: Apis mellifera (honeybee)

C/Date: 19-Feb-1984 #sequence\_revision 19-Feb-1984 #text\_change 15-Sep-2000

C/Accession: A91133; A91640; B61285; S23131; A01761; A18860

R/Vlasak, R.; Unger-Ullmann, C.; Kreil, G.; Fritsch, A.M.

Eur. J. Biochem. 135, 123-126, 1983

A/Title: Nucleotide sequence of cloned cDNA coding for honeybee prepro-melittin.

A/Reference number: A91133; MUID:83287387; PMID:6309516

A/Accession: A91133

A/Molecule type: mRNA

A/Residues: 1-70 &lt;VLA&gt;

A/Cross-references: GB:X02007; NID:95621; PIDN:CA826038.1; PID:95622

R/Habermann, E.; Jenesch, J.

Hoppe-Seyler's Z. Physiol. Chem. 348, 37-50, 1967

A/Title: Sequenzanalyse des Melittins aus den tryptischen und peptischen Spaltstuecken.

A/Reference number: A91640; MUID:68327913; PMID:5592400

A/Accession: A91640

A/Molecule type: protein

A/Residues: 44-69 &lt;HAB&gt;

R/Gaudier, J.; Hanson, J.M.; Shiplini, R.A.; Vernon, C.A.

Eur. J. Biochem. 83, 405-410, 1978

A/Title: The structures of some peptides from bee venom.

A/Reference number: A91253; MUID:78126668; PMID:631126

A/Accession: B61285

A/Molecule type: protein

A/Residues: 51-69 &lt;GAU&gt;

R/Ramalingam, K.; Bello, J.

Biochem. J. 284, 663-665, 1992

A/Title: Effect of permethylation on the haemolytic activity of melittin.

A/Reference number: S23131; MUID:92321983; PMID:1622887

A/Accession: S23131

A/Status: preliminary

A/Molecule type: protein

A/Residues: 44-69 &lt;RAM&gt;

R/Schroeder, E.; Luebke, K.; Lehmann, M.; Beetz, I.

Experientia 27, 764-765, 1971

A/Title: Haemolytic activity and action on the surface tension of aqueous solutions of s

A/Reference number: A91266; MUID:72098668; PMID:5139482

A/Contents: annotation; synthesis

A/Note: the structure of melittin was confirmed by synthesis of a peptide with full hemo

R/Kreil, G.; Kreil-Kiss, G.

Biochem. Biophys. Res. Commun. 27, 275-280, 1967

A/Title: The isolation of N-formylglycine from a polypeptide present in bee venom.

A/Reference number: A90165; MUID:67248282; PMID:6040373

A/Contents: annotation

A/Note: about 10% of melittin is formylated at the amino end

R/Luebke, K.; Matthes, S.; Kloss, G.

Experientia 27, 765-767, 1971

A/Title: Isolation and structure of N(alpha)-formyl melittin.

A/Reference number: A91267; MUID:72098669; PMID:5139483

A/Contents: annotation; synthesis

A/Note: N-formyl-melittin major was isolated from the venom and its structure was confir

R/Mollay, C.; Vlasak, U.; Kreil, G.

Proc. Natl. Acad. Sci. U.S.A. 79, 2260-2263, 1982

A/Title: Cleavage of honeybee prepro-melittin by an endoprotease from rat liver microsomes

A/Reference number: A18880; MUID:82247982; PMID:7048315

A/Contents: annotation; intact signal sequence after cleavage by partially purified sign

R/Eisenberg, D.; Griboskov, M.; Terwilliger, T.C.

submitted to the Brookhaven Protein Data Bank, October 1990

A/Reference number: A50496; PDB:2MLT

A/Contents: annotation; X-ray crystallography, 2.0 angstroms, residues 44-69

R/Terwilliger, T.C.; Eisenberg, D.

J. Biol. Chem. 257, 6010-6015, 1982

A/Title: The structure of melittin. I. Structure determination and partial refinement.

A/Reference number: A30639; MUID:82189958; PMID:7076661

A/Contents: annotation; X-ray crystallography, 2.0 angstroms

R/Terwilliger, T.C.; Eisenberg, D.

J. Biol. Chem. 257, 6016-6022, 1982

A/Title: The structure of melittin. II. Interpretation of the structure.

A/Reference number: A30640; MUID:82189959; PMID:7076662

A/Contents: annotation; X-ray crystallography, 2.0 angstroms

C/Superfamily: melittin major

C/Keywords: amidated carboxyl end; blocked amino end; hemolysis; homotetramer; venom

F/1-21/Domain: signal sequence #status experimental &lt;SIG&gt;

F/22-43/Domain: propeptide #status experimental &lt;PRO&gt;

F/44-69/Product: melittin F #status experimental &lt;MAU&gt;

F/51-66/Product: melittin F #status experimental &lt;MBU&gt;

F/44/Modified site: formylated amino end (Gly) (in mature form) (partial) #status exper

F/69/Modified site: amidated carboxyl end (Gln) (amide in mature form from following glyc

Query Match 10.8%; Score 102; DB 1; Length 70;

Best Local Similarity 100.0%; Pred. No. 0.0047;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKFLVNVAVLVVVVVISYIYA 21

Db 1 MKFLVNVAVLVVVVVISYIYA 21

```
RESULT 4
VGBED2
glycoprotein D precursor - human herpesvirus 1 (strain Hzt)
C/Species: human herpesvirus 1
C/Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #ext_change 09-Sep-1994
C/Accession: A90945; A03729
R/Lasky, L.A.; Dowbenko, D.J.
DNA 3, 23-29, 1984
A/Title: DNA sequence analysis of the type-common glycoprotein-D genes of herpes simplex
A/Reference number: A90945; WUID:84131549; PMID:6321120
A/Accession: A90945
A/Molecule type: DNA
A/Residues: 1-393 <LAS>
A/Cross-references: GB:K02372
C/Superfamily: herpesvirus glycoprotein D
C/Keywords: glycoprotein; transmembrane protein
F:1-20/Domain: signal sequence #status predicted <SIG>
F:21-393/Product: glycoprotein D #status predicted <GPD>
F:141-360/Domain: transmembrane #status predicted <TMN>
F:118,145,286/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match          10.5%; Score 99; DB 1; Length 393;
Best Local Similarity 100.0%; Pred. No. 0.076;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 159 OPELAPEDPEDSALLDPV 177
DB 289 OPELAPEDPEDSALLDPV 307

RESULT 5
VGBED1
glycoprotein D - human herpesvirus 1
C/Species: human herpesvirus 1
C/Date: 05-Apr-1983 #sequence_revision 05-Apr-1983 #ext_change 16-Jul-1999
C/Accession: A94268; B90945; A03729
R/Watson, R.J.; Weis, J.H.; Salstrom, J.S.; Enquist, L.W.
Science 218, 381-384, 1982
A/Title: Herpes simplex virus type-1 glycoprotein D gene: nucleotide sequence and expres
A/Reference number: A94268; WUID:83016330; PMID:6289440
A/Accession: A94268
A/Molecule type: DNA
A/Residues: 1-394 <MAT>
A/Cross-references: GB:J02217; NID:G330100; PIN:AAA5785.1; PID:G330101
A/Experimental source: strain Patton
A/Note: a strongly hydrophobic region of 25 amino acids between residues 340 and 364 is
R/Lasky, L.A.; Dowbenko, D.J.
DNA 3, 23-29, 1984
A/Title: DNA sequence analysis of the type-common glycoprotein-D genes of herpes simplex
A/Reference number: A90945; WUID:84131549; PMID:6321120
A/Accession: B90945
A/Molecule type: DNA
A/Residues: 1-3, 'A', 'S', '70', 'N', '72-83, 85-269, 'R', '271-282, 'P', '284-364, 'R', '366-394 <LAS>
A/Experimental source: strain Hzt
C/Superfamily: herpesvirus glycoprotein D
C/Keywords: glycoprotein; transmembrane protein
F:119,146,287/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match          10.5%; Score 99; DB 1; Length 394;
Best Local Similarity 100.0%; Pred. No. 0.076;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 159 OPELAPEDPEDSALLDPV 177
DB 290 OPELAPEDPEDSALLDPV 308

RESULT 6
VGBE17
glycoprotein D precursor - human herpesvirus 1 (strain 17)
C/Species: human herpesvirus 1
```

```
A/Note: host Homo sapiens (man)
C/Date: 17-Mar-1987 #sequence_revision 17-Mar-1987 #ext_change 16-Jun-2000
C/Accession: A03730
R/McGeoch, D.J.; Dolan, A.; Donald, S.; Rixon, F.J.
J. Mol. Biol. 181, 1-13, 1985
A/Title: Sequence determination and genetic content of the short unique region in the ger
A/Reference number: A00656; WUID:85160822; PMID:2984429
A/Accession: A03730
A/Molecule type: DNA
A/Residues: 1-394 <MCG>
A/Cross-references: GB:X14112; NID:G1944536; PIN:CAA32283.1; PID:G59564
C/Superfamily: herpesvirus glycoprotein D
C/Keywords: glycoprotein; transmembrane protein
F:1-20/Domain: signal sequence #status predicted <SIG>
F:21-394/Product: glycoprotein D #status predicted <GPD>
F:342-358/Domain: transmembrane #status predicted <TMN>
F:119,146,287/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match          10.5%; Score 99; DB 1; Length 394;
Best Local Similarity 100.0%; Pred. No. 0.076;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 159 OPELAPEDPEDSALLDPV 177
DB 290 OPELAPEDPEDSALLDPV 308

RESULT 7
A47627
glycoprotein D precursor - human herpesvirus 1 (strain ANG)
C/Species: human herpesvirus 1
C/Date: 17-Feb-1994 #sequence_revision 17-Feb-1994 #ext_change 16-Jul-1999
C/Accession: A47627
R/Izumi, K.M.; Stevens, J.G.
J. Exp. Med. 172, 487-496, 1990
A/Title: Molecular and biological characterization of a herpes simplex virus type 1 (HSV-
A/Reference number: A47627; WUID:90324869; PMID:2165127
A/Accession: A47627
A/Molecule type: DNA
A/Residues: 1-394 <IZU>
A/Cross-references: GB:X53361; NID:G60414; PIN:CAA38245.1; PID:G60415
C/Superfamily: herpesvirus glycoprotein D
C/Keywords: glycoprotein; transmembrane protein
F:1-20/Domain: signal sequence #status predicted <SIG>
F:21-394/Product: glycoprotein D #status predicted <GPD>
F:342-360/Domain: transmembrane #status predicted <TMN>
F:119,146,287/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match          10.5%; Score 99; DB 1; Length 394;
Best Local Similarity 100.0%; Pred. No. 0.076;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 159 OPELAPEDPEDSALLDPV 177
DB 290 OPELAPEDPEDSALLDPV 308

RESULT 8
OFFFCP
copia polypeptide - fruit fly (Drosophila melanogaster) retrotransposon copia
C/Species: copia protein, 31k; copia protein, 48k; proteinase
C/Date: 04-Dec-1986 #sequence_revision 04-Dec-1986 #ext_change 16-Jul-1999
C/Accession: A03324; S03612; S14635
R/Mount, S.M.; Rubin, G.M.
Mol. Cell. Biol. 5, 1630-1638, 1985
A/Title: Complete nucleotide sequence of the Drosophila transposable element copia: homo1
A/Reference number: A03324; WUID:85267679; PMID:2410772
A/Accession: A03324
A/Molecule type: DNA
A/Residues: 1-1409 <MOU>
A/Cross-references: GB:M1240; NID:G158615; PIN:AAA74497.1; PID:G950318
R/Miller, K.; Rosenbaum, J.; Zbrzezna, V.; Pogo, A.O.
```







GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 22, 2004, 11:37:25 ; Search time 58.5973 Seconds  
(without alignments)  
779.478 Million cell updates/sec

Title: US-09-890-806-5  
Perfect score: 947  
Sequence: 1 MFLVVALVFMVYISYIY.....GQELAPDEPDSDALLEPVPV 177

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues  
Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

1: SP archaea: \*  
2: SP bacteria: \*  
3: SP fungi: \*  
4: SP human: \*  
5: SP invertebrate: \*  
6: SP mammal: \*  
7: SP mhc: \*  
8: SP organelle: \*  
9: SP phage: \*  
10: SP plant: \*  
11: SP rodent: \*  
12: SP virus: \*  
13: SP vertebrate: \*  
14: SP unclassified: \*  
15: SP virus: \*  
16: SP bacteriap: \*  
17: SP archaeap: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	489	51.6	203	12 Q8QRV5	Q8QRV5 chimpanzee
2	136	14.4	186	12 Q918F3	Q918F3 human cytom
3	133	14.0	186	12 Q918F0	Q918F0 human cytom
4	132	13.9	186	12 Q910T7	Q910T7 human cytom
5	131	13.8	186	12 Q918E4	Q918E4 human cytom
6	130	13.7	186	12 Q918E6	Q918E6 human cytom
7	127	13.4	196	12 Q8UKF7	Q8UKF7 cercopithec
8	124	13.1	186	12 Q918E2	Q918E2 human cytom
9	121	12.8	186	12 Q918E9	Q918E9 human cytom
10	118	12.5	186	12 Q910G4	Q910G4 human cytom
11	115	12.1	186	12 Q8UKZ6	Q8UKZ6 human cytom
12	112.5	11.9	186	12 Q918F4	Q918F4 human cytom
13	109.5	11.6	187	12 Q8QRV4	Q8QRV4 chimpanzee
14	108.5	11.5	149	12 Q910V7	Q910V7 human cytom
15	107.5	11.4	149	12 Q918F1	Q918F1 human cytom
16	106.5	11.2	149	12 Q918E7	Q918E7 human cytom

17	101.5	10.7	149	12 Q69186	Q69186 human cytom
18	100.5	10.6	149	12 Q918E3	Q918E3 human cytom
19	99	10.5	394	12 Q8QRB7	Q8QRB7 human cytom
20	99	10.5	394	12 Q918W3	Q918W3 human cytom
21	99	10.5	394	12 Q69081	Q69081 human cytom
22	99	10.5	394	12 Q05059	Q05059 human cytom
23	99	10.5	394	12 Q05060	Q05060 human cytom
24	99	10.5	394	12 Q8QRB6	Q8QRB6 human cytom
25	99	10.5	394	12 Q69082	Q69082 human cytom
26	97.5	10.3	149	12 Q910D0	Q910D0 human cytom
27	96.5	10.2	1017	5 Q8T391	Q8T391 drosophila
28	91	9.6	393	12 Q69467	Q69467 herpes simp
29	91	9.6	394	12 Q918H6	Q918H6 human cytom
30	85	9.0	376	16 Q98FC4	Q98FC4 rhizobium 1
31	84.5	8.9	1409	5 Q08461	Q08461 drosophila
32	84.5	8.9	1578	16 Q92E25	Q92E25 listeria in
33	84	8.9	159	17 Q91Y00	Q91Y00 pyrococcus
34	84	8.9	254	16 Q8ZF06	Q8ZF06 yersinia pe
35	83	8.8	318	16 Q93J07	Q93J07 streptomyce
36	82	8.7	1938	2 P70983	P70983 bacillus sp
37	80.5	8.5	404	16 Q8CVG5	Q8CVG5 escherichia
38	80.5	8.5	415	5 Q19894	Q19894 caenorhabdi
39	80.5	8.5	1582	16 Q8Y9A5	Q8Y9A5 listeria mo
40	80	8.4	250	16 Q8BX2	Q8BX2 xanthomonas
41	80	8.4	725	2 Q59239	Q59239 bacillus sp
42	79.5	8.4	692	2 Q30565	Q30565 bacillus br
43	79	8.3	159	17 Q59596	Q59596 pyrococcus
44	79	8.3	325	16 Q8UJ00	Q8UJ00 agrobacteri
45	79	8.3	329	4 Q8N225	Q8N225 homo sapien

#### ALIGNMENTS

RESULT 1  
Q8QRV5 PRELIMINARY; PRT; 203 AA.  
AC Q8QRV5; 08QRV5; 01-JUN-2002 (TREMREL. 21, Created)  
DT 01-JUN-2002 (TREMREL. 21, Last sequence update)  
DT 01-JUN-2002 (TREMREL. 21, Last annotation update)  
DE US2.  
OS Chimpanzee cytomegalovirus.  
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
OC Betaherpesvirinae; Cytomegalovirus.  
OX NCBI\_TaxID=188763;  
RN (1)  
RP SEQUENCE FROM N.A.  
RA Davidson A.J., Akter P., Dolan A., Wright K.M., Addison C.,  
RA Alencador D.J., Hayward G.S., McGeoch D.J.;  
RT "The human cytomegalovirus genome revisited";  
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF480884; AAM00783.1; -  
SQ SEQUENCE 203 AA; 23262 MW; CA131FD182BE8300 CRC64;

Query Match 51.6%; Score 489; DB 12; Length 203;  
Best Local Similarity 63.4%; Pred. No. 6.2e-41;

Matches 85; Conservative 24; Mismatches 23; Indels 2; Gaps 1;

QY	24	PDGITKAGDALARPMKSTAKHPMFOIEDNRKYIDNGKLFARQSIYGNMRFYFDPKADYG	83
DB	28	PDPHTEI--TSYKPMKSTATRPMTIDENRCHIENGQMGSGVSGNLTFVDFDPKADYG	85
QY	84	GIQENLYHADVEFVPGESLKNVYRNLDMVPIFETLALRYLQGVYLRKYPELRUY	143
DB	86	GIQENLYHADVEFVPGESLKNVYRNLDMVPIFETLALRYLQGVYLRKYPELRUY	145
QY	144	TSSAYVMNNQYGVY	157
DB	146	TSAYVMNNQYGVY	159

RESULT 2





Matches	41.	Conservative	33.	Mismatches	52.	Indels	42.	Gaps	9.
Qy	10	VPMVVYISYIYARLPDGIITKAGEDALRPWK-----	STAKHPFQIEDNRCYIDNGKL	PAR	64				
Db	4	VLVLAIALVLFLRLADSV-----	FWPIIDVVVSEIRSAHFVEENQCWFHGM	LYEK	54				
Qy	65	GSIVGNMRFVDPADYGVGENL-----	YHADVVEFVPGESL	KKNV-----R	109				
Db	55	GRMSGNFEFEKH-----	VNVGISQSYMDRLQVSGEQYHDE---	RGAYFENNIGHPVTH	107				
Qy	110	NLDVPIPETLTALRLVLOGD-VIMRLCVPBELRVDYTS	SAYMMNQYGM	156					
Db	108	TVDWVDI--TLSTR---	WGDPKRYACVPGVRMDYSSQTINWYLQ	RSKM	150				

  

RESULT 6									
ID	Q918B6	PRELIMINARY;		PRT;	186	AA.			
AC	O918E6								
DT	01-DEC-2001	(TREMBLrel. 19, Created)							
DT	01-DEC-2001	(TREMBLrel. 19, Last sequence update)							
DT	01-DEC-2001	(TREMBLrel. 19, Last annotation update)							
DE	US3 protein.								
GN	US3.								
OS	Human cytomegalovirus.								
OC	Viruses; dsDNA viruses, no RNA stage; Herpesviridae;								
OC	Betaherpesvirinae; Cytomegalovirus.								
OX	NCBI_TaxID=10359;								
RN	[1]								
RP	SEQUENCE FROM N.A.								
RC	STRAIN=9A;								
RA	Scott G.M., Bartell B.G., Oram J., Rawlinson W.D.;								
RT	"Characterization of transcripts from the human cytomegalovirus genes								
RL	TRL7, UL20a, UL36, UL65, UL94, US3 and US34."								
RL	Submitted (Aug-2001) to the EMBL/Genbank/DBJ databases.								
SO	EMBL: AF413657; AAL14501.1; ..	7C0318DABA6E91D	CRC64;						

  

Query Match	13.7%;	Score 130;	DB 12;	Length 186;
Best Local Similarity	24.5%;	Pred. No. 4.9e-05;		
Matches	40;	Conservative	36;	Mismatches 55; Indels 32; Gaps 8;
Qy	10	VPMVVYISYIYARLPDGIITKAGEDALRPWKSTAKHPFQIEDNRCYIDNGKL	PAR	GISTVG 69
Db	4	VMMALIALVLFLRLADSVPRPLDVV----	SEIRSAHFVEENQCWFHGM	LYEKGKRS 59
Qy	70	NMSRVPDPKADYGVGENL-----	YHADVVEFVPGESL	KKNV-----RNLDMV 114
Db	60	NFTEKGF--VNVGISQSYMDRLQVSGEQYHDE---	RGAYFENNIGHPVTH	TVDMV 112
Qy	115	PIFETLALRLVLOGD-VIMRLCVPBELRVDYTS	SAYMMNQYGM	156
Db	113	DI--TLSTR---	WGDPKRYACVPGVRMDYSSQTINWYLQ	RSKM 150

  

RESULT 7									
ID	O8JKE7	PRELIMINARY;		PRT;	196	AA.			
AC	O8JKE7								
DT	01-OCT-2002	(TREMBLrel. 22, Created)							
DT	01-OCT-2002	(TREMBLrel. 22, Last sequence update)							
DT	01-MAR-2003	(TREMBLrel. 23, Last annotation update)							
DE	US2 protein.								
GN	US2.								
OS	Cercopithecine herpesvirus 8.								
OC	Viruses; dsDNA viruses, no RNA stage; Herpesviridae;								
OC	Betaherpesvirinae; Cytomegalovirus.								
OX	NCBI_TaxID=47929;								
RN	[1]								
RP	SEQUENCE FROM N.A.								
RX	MEDLINE=22174937; PubMed=12186931;								
RA	Chang W.L., Tataratal A.F., Zhou S.S., Borowsky A.D., Barry P.A.;								
RA	"A recombinant threus cytomegalovirus expressing enhanced green								

Query Match	13.4%	Score 127	DB 12	Length 196
Best Local Similarity	26.1%	Pred. No. 0.0001		
Matches	29	Conservative	28	Mismatches 50; Indels 4; Gaps 4
Qy	42	AKHPWFOEDNRRCYINDGKLFARSGIVNMSRFVDPKADYGVG-ENLYVHADVEFVP	100	
Db	37	ARRALYVYQACVACLEGKLFMTGTYITINIDSY-FVQVYVYTRDKIEDLQPFKQMK-LS	94	
Qy	101	GESLKMVNRINDVMPPIFETLALRLVL-QGDVIMRCVPELRVDYTSAYMM	150	
Db	95	SDKLEVDLRYEIDWTTTKVEMRSLSNMSDIIMWTCBPHVKPDPFLTHNYLW	145	
RESULT 8				
ID	Q918E2	PRELIMINARY;	PRT;	186 AA.
AC	Q918E2;			
DT	01-DEC-2001 (Tremblrel. 19, Created)			
DT	01-DEC-2001 (Tremblrel. 19, Last sequence update)			
DT	01-DEC-2001 (Tremblrel. 19, Last annotation update)			
GN	US3 protein.			
OS	Human cytomegalovirus.			
OC	Virusae; dsDNA virusae, no RNA stage; Herpesviridae;			
OC	Betaherpesvirinae; Cytomegalovirus.			
NCBI_TaxID=10359;				
RN	SEQUENCE FROM N.A.			
RC	STRAIN=95A.			
RA	Scott G.M., Bartell B.G., Oram J., Rawlinson W.D.;			
RT	"Characterization of transcripts from the human cytomegalovirus genes			
RT	TRL7, UL20a, UL36, UL65, UL94, US3 and US34."			
RL	Submitted (AUG-2001) to the EMBL/Genbank/DBJ databases.			
DR	EMBL; AF413665; AAL14524.1; -			
SD	SEQUENCE 186 AA; 21602 MW; 3BB0504768183511 CRC64;			
Query Match	13.1%	Score 124	DB 12	Length 186;
Best Local Similarity	23.9%	Pred. No. 0.00019;		
Matches	39	Conservative	36	Mismatches 56; Indels 32; Gaps 8
Qy	10	VFMVVVISYIARLPDGTITKAGEDALRWKSTAKHPWFOEDNRCYINDGKLFARGSIYV	69	
Db	4	VLVAITLAVLFLRLADSVPRPLDVV---SEIRSAHFRVBEHQCFHMGMLYFKGRMSG	59	
Qy	70	NMSRFVDPKADYGVGENTL-----YHADVEFVPEESLKMV-----NLDVM	114	
Db	60	NFETRRH---VNGIVSKSYMDRLQVSGEQHHDE---RQAYFPMNIGGHHVTHTVDMV	112	
Qy	115	PIFETLALRLVLQGD-VIMRCVPELRVDYTSAYMMNQXM	156	
Db	113	DI--TLSTR---WGDPPKYACACQVPRMDYSSQITINWYLQKSM	150	
RESULT 9				
ID	Q918E9	PRELIMINARY;	PRT;	186 AA.
AC	Q918E9;			
DT	01-DEC-2001 (Tremblrel. 19, Created)			
DT	01-DEC-2001 (Tremblrel. 19, Last sequence update)			
DT	01-DEC-2001 (Tremblrel. 19, Last annotation update)			
GN	US3 protein.			
OS	Human cytomegalovirus.			
OC	Virusae; dsDNA virusae, no RNA stage; Herpesviridae;			
OC	Betaherpesvirinae; Cytomegalovirus.			
NCBI_TaxID=10359;				
RN	[1]			

RP SEQUENCE FROM N.A.  
 RC STRAIN=27A;  
 RA Scott G.M., Barrell B.G., Oram J., Rawlinson W.D.;  
 RT "Characterization of transcripts from the human cytomegalovirus genes  
 TRL7, UL20a, UL56, UL94, US3 and US34."  
 RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF413651; AAL14483.1; -  
 SO SEQUENCE 186 AA; 2153 MW; CB68FE1B6C76728 CRC64;  
 Query Match 12.8%; Score 121; DB 12; Length 186;  
 Best Local Similarity 24.1%; Pred. No. 0.00039;  
 Matches 39; Conservative 32; Mismatches 62; Indels 28; Gaps 7;  
 QY 10 VFMVVVSYIYARLPDGTGAGEDALRPWKSTAGHPFQIEDNRCYIDNGKLFARGSIYV 69  
 DB 4 VLVATIAVLFLRLADSVPRPLDVV---SEIRSAHFVEENQCFHMGMLHYKGRMSG 59  
 QY 70 NMSR-----FVFDPKADYGVGENLYHADDFEVPESLKNV-----RNLDVMP 116  
 DB 60 NFEKHFVSVGVISQSYMDRLQVSGEQYHDE-----RGAYFEMNIGHPVPHVTVDV 114  
 QY 117 FETLALRLVLQGD-VIMLRQVPELRVDYTSAYMMNQYGM 156  
 DB 115 --TLSTR--WGDPKKYAACVQVRMDYSSQITINWYLQISM 150  
 RESULT 10  
 Q910G4 PRELIMINARY; PRT; 186 AA.  
 ID Q910G4;  
 AC Q910G4;  
 DT 01-DEC-2001 (T-EMBLrel. 19, Created)  
 DT 01-DEC-2001 (T-EMBLrel. 19, Last sequence update)  
 DT 01-MAR-2002 (T-EMBLrel. 20, Last annotation update)  
 DE US3 protein.  
 GN US3.  
 OS Human cytomegalovirus (strain Towne), and  
 OS Human cytomegalovirus.  
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
 OC Betaherpesvirinae; Cytomegalovirus.  
 NC NCB1\_taxid=10363, 10359;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=Human cytomegalovirus; STRAIN=29A;  
 RA Scott G.M., Barrell B.G., Oram J., Rawlinson W.D.;  
 RT "Characterization of transcripts from the human cytomegalovirus genes  
 TRL7, UL20a, UL56, UL94, US3 and US34."  
 RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AY038933; AAK11882.1; -  
 SO SEQUENCE 186 AA; 2151 MW; CB68FE1B6D6B0D8 CRC64;  
 Query Match 12.5%; Score 118; DB 12; Length 186;  
 Best Local Similarity 24.1%; Pred. No. 0.00078;  
 Matches 38; Conservative 32; Mismatches 60; Indels 28; Gaps 7;  
 QY 10 VFMVVVSYIYARLPDGTGAGEDALRPWKSTAGHPFQIEDNRCYIDNGKLFARGSIYV 69  
 DB 4 VLVATIAVLFLRLADSVPRPLDVV---SEIRSAHFVEENQCFHMGMLHYKGRMSG 59  
 QY 70 NMSR-----FVFDPKADYGVGENLYHADDFEVPESLKNV-----RNLDVMP 116  
 DB 60 NFEKHFVSVGVISQSYMDRLQVSGEQYHDE-----RGAYFEMNIGHPVPHVTVDV 114  
 QY 117 FETLALRLVLQGD-VIMLRQVPELRVDYTSAYMMNQYGM 153  
 DB 115 --TLSTR--WGDPKKYAACVQVRMDYSSQITINWYLQ 147

RESULT 11  
 Q8UZK6 PRELIMINARY; PRT; 186 AA.  
 ID Q8UZK6;  
 AC Q8UZK6;  
 DT 01-MAR-2002 (T-EMBLrel. 20, Created)  
 DT 01-MAR-2002 (T-EMBLrel. 20, Last sequence update)  
 DT 01-MAR-2002 (T-EMBLrel. 20, Last annotation update)  
 DE US3 protein.  
 OS Human cytomegalovirus (strain Towne).  
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
 OC Betaherpesvirinae; Cytomegalovirus.  
 NC NCB1\_taxid=10363;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Towne;  
 RA "Immune evasion genes from the Towne strain of human  
 cytomegalovirus."  
 RT Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AY072774; AAL67142.1; -  
 SO SEQUENCE 186 AA; 21335 MW; 86D0167FD66BAB84 CRC64;  
 Query Match 12.1%; Score 115; DB 12; Length 186;  
 Best Local Similarity 24.1%; Pred. No. 0.0015;  
 Matches 38; Conservative 32; Mismatches 60; Indels 28; Gaps 7;  
 QY 10 VFMVVVSYIYARLPDGTGAGEDALRPWKSTAGHPFQIEDNRCYIDNGKLFARGSIYV 69  
 DB 4 VLVATIAVLFLRLADSVPRPLDVV---SEIRSAHFVEENQCFHMGMLHYKGRMSG 59  
 QY 70 NMSR-----FVFDPKADYGVGENLYHADDFEVPESLKNV-----RNLDVMP 116  
 DB 60 NFEKHFVSVGVISQSYMDRLQVSGEQYHDE-----RGAYFEMNIGHPVPHVTVDV 114  
 QY 117 FETLALRLVLQGD-VIMLRQVPELRVDYTSAYMMNQYGM 153  
 DB 115 --TLSTR--WGDPKKYAACVQVRMDYSSQITINWYLQ 147  
 RESULT 12  
 Q918F4 PRELIMINARY; PRT; 149 AA.  
 ID Q918F4;  
 AC Q918F4;  
 DT 01-DEC-2001 (T-EMBLrel. 19, Created)  
 DT 01-DEC-2001 (T-EMBLrel. 19, Last sequence update)  
 DT 01-DEC-2001 (T-EMBLrel. 19, Last annotation update)  
 DE US3 protein.  
 GN US3.  
 OS Human cytomegalovirus.  
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
 OC Betaherpesvirinae; Cytomegalovirus.  
 NC NCB1\_taxid=10359;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=22A;  
 RA Scott G.M., Barrell B.G., Oram J., Rawlinson W.D.;  
 RT "Characterization of transcripts from the human cytomegalovirus genes  
 TRL7, UL20a, UL56, UL94, US3 and US34."  
 RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF413648; AAL14475.1; -  
 SO SEQUENCE 149 AA; 17217 MW; E822385AC2CE5420 CRC64;  
 Query Match 11.9%; Score 112.5; DB 12; Length 149;  
 Best Local Similarity 25.4%; Pred. No. 0.0021;  
 Matches 43; Conservative 32; Mismatches 49; Indels 45; Gaps 11;  
 QY 10 VFMVVVSYIYARLPDGTGAGEDALRPWKSTAGHPFQIEDNRCYIDNGKLFARGSIYV 69  
 DB 4 VLVATIAVLFLRLADSVPR---PLNVVSEIKSAHFVEENQCFHMGMLHYKGRMSG 59  
 QY 70 NMSRFVDPKADYGVG---ENLYV---HADDFEVPESLKNV-----RNL 111









Db 208 YEDSIYRNLYDADYDNTVMDQYKESIKFWLDKIDGIRD 251

RESULT 4

ZNUB\_ECOLI STANDARD; PRT; 261 AA.  
AC P39832; P76286;  
DT 01-FEB-1995 (Rel. 31, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE High-affinity zinc uptake system membrane protein znuB.  
GN ZNUB OR B1859.  
OS Escherichia coli.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Escherichia.  
NCBI\_TaxId=562;  
[1]  
RN SEQUENCE FROM N.A.  
RC STRAIN=K12 / MG1655;  
RX MEDLINE=97426617; PubMed=9278503;  
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
RA Mau B., Shao Y.;  
RT "The complete genome sequence of Escherichia coli K-12.";  
RL Science 277:1453-1474(1997).  
[2]  
RN SEQUENCE FROM N.A.  
RC STRAIN=K12;  
RX MEDLINE=97251358; PubMed=9097040;  
RA Itoh T., Aiba H., Baba T., Fujita K., Hayashi K., Inada T.,  
RA Isono K., Kasai H., Kimura S., Kitakawa M., Kitagawa M.,  
RA Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,  
RA Nakade S., Nakamura Y., Nishimoto H., Nishio Y., Oshima T.,  
RA Saito N., Sampei G., Seki Y., Sivasubramanian S., Tagami H.,  
RA Takeda J., Takemoto K., Wada C., Yamamoto Y., Horiiuchi T.;  
RT "A 460-kb DNA sequence of the Escherichia coli K-12 genome  
corresponding to the 40.1-50.0 min region on the linkage map.";  
RL DNA Res. 3:379-392(1996).  
[3]  
RN SEQUENCE OF 1-92 FROM N.A.  
RC STRAIN=K12 / EMG2;  
RA Robison K., O'Keefe T., Church G.M.;  
RN Submitted (OCT-1995) to the EMBL/GenBank/DBJ databases.  
[4]  
RN SEQUENCE OF 80-261 FROM N.A.  
RX MEDLINE=88314937; PubMed=2842314;  
RA Shinagawa H., Makino K., Amemura M., Kimura S., Iwaseki H., Nakata A.;  
RT "Structure and regulation of the Escherichia coli ruv operon involved  
in DNA repair and recombination.";  
RL J. Bacteriol. 170:4322-4329(1988).  
[5]  
RN IDENTIFICATION.  
RX MEDLINE=95075659; PubMed=7984428;  
RA Borodovsky M., Rudd K.E., Koonin E.V.;  
RT "Intrinsic and extrinsic approaches for detecting genes in a  
bacterial genome.";  
RL Nucleic Acids Res. 22:4756-4767(1994).  
[6]  
RN CHARACTERIZATION.  
RX MEDLINE=98343803; PubMed=9680209;  
RA Patzer S.I., Hantke K.;  
RT "The znuABC high-affinity zinc uptake system and its regulator Zur in  
Escherichia coli.";  
RL Mol. Microbiol. 28:1199-1210(1998).  
CC -1- FUNCTION: INVOLVED IN THE HIGH-AFFINITY ZINC UPTAKE TRANSPORT  
SYSTEM.  
CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane  
(Probable).  
CC -1- SIMILARITY: BELONGS TO THE ABC-3 SUBFAMILY OF INTEGRAL MEMBRANE  
PROTEINS. STRONG, TO H. INFLUENZAE ZNUB.  
-----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----

DR EMBL; AE000280; AAC74929.1; -.  
DR EMBL; D90828; BAA15657.1; -.  
DR EMBL; D90829; BAA15670.1; -.  
DR EMBL; U38702; AAA81031.1; -.  
DR EMBL; M21298; -; NOT\_ANNOTATED\_CDS.  
DR PIR; C64948; C64948.  
DR Ecogene; EG12368; znuB.  
DR InterPro; IPR001626; ABC\_transp3.  
DR Pfam; PF00950; ABC-3; 1.  
KW Transport; Zinc transport; Transmembrane; Inner membrane;  
KW Complete proteome.  
FT TRANSMEM 8 28 POTENTIAL.  
FT TRANSMEM 54 74 POTENTIAL.  
FT TRANSMEM 84 104 POTENTIAL.  
FT TRANSMEM 122 142 POTENTIAL.  
FT TRANSMEM 178 198 POTENTIAL.  
FT TRANSMEM 214 234 POTENTIAL.  
FT TRANSMEM 236 256 POTENTIAL.  
SQ SEQUENCE 261 AA; 27728 MW; 78382B2EACE1490 CRC64;

Query Match 11.8%; Score 74; DB 1; Length 261;  
Best Local Similarity 40.6%; Pred. No. 1.1;  
Matches 13; Conservative 8; Mismatches 11; Indels 0; Gaps 0;

OY 82 LDVMPFETLRLVQLQGDVWLRCPBELRD 113  
Db 52 LDVNPYAVIATVTLHLAGLWLEKRPOLAD 83

RESULT 5

ID COP1\_DROME STANDARD; PRT; 1409 AA.  
AC P04146; Q03728; Q24280; Q24555; Q24585; Q24586; Q24587;  
DT 01-NOV-1986 (Rel. 03, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Copia protein [Contains: Copia VLP protein; Copia protease  
(EC 3.4.23.-)].  
GN COP1.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxId=7227;  
RN (1)  
RN SEQUENCE FROM N.A. (ISOFORM LONG).  
RX MEDLINE=85267679; PubMed=2410772;  
RA Mount S.M., Rubin G.M.;  
RT "Complete nucleotide sequence of the Drosophila transposable element  
copia: homology between copia and retroviral proteins.";  
RL Mol. Cell. Biol. 5:1630-1638(1985).  
[2]  
RN SEQUENCE FROM N.A., SEQUENCE OF 2-10, AND ALTERNATIVE SPLICING.  
RX MEDLINE=85540569; PubMed=2409449;  
RA Emori Y., Shiba T., Kanaya S., Inouye S., Yuki S., Saigo K.;  
RT "The nucleotide sequences of copia and copia-related RNA in Drosophila  
virus-like particles.";  
RL Nature 315:773-776(1985).  
[3]  
RN SEQUENCE FROM N.A. (ISOFORM SHORT).  
RX MEDLINE=89183629; PubMed=2538806;  
RA Miller K., Rosenbaum J., Zbrzezna V., Pogo A.O.;  
RT "The nucleotide sequence of Drosophila melanogaster copia-specific  
2.1-kb mRNA.";  
RL Nucleic Acids Res. 17:2134-2134(1989).  
-----

[4]  
 RN SEQUENCE FROM N.A. (ISOFORM SHORT), AND MUTAGENESIS OF ASP-292.  
 RP TISSUE-LARVA; PubMed=1689241;  
 RX MEDLINE=90151630; PubMed=1689241;  
 RA Yoshiooka K., Honma H., Zushi M., Kondo S., Togaishi S., Miyake T.,  
 RA Shiba T.;  
 RT "virus-like particle formation of Drosophila copia through  
 RT autocatalytic processing."; EMBO J. 9:535-541(1990).  
 CC -1- ALTERNATIVE PRODUCTS;  
 CC Event=Alternative splicing; Named isoforms=2;  
 CC Name=Long;  
 CC IsoId=P04146-1; Sequence=Displayed;  
 CC Name=Short;  
 CC IsoId=P04146-2; Sequence=VSP\_005226;  
 CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY A11.  
 CC -1- SIMILARITY: Contains 1 CCHC-type zinc finger.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; X04456; CAA28054.2; -  
 DR EMBL; X02599; CAA26444.1; -  
 DR EMBL; X02599; CAA26445.1; -  
 DR EMBL; X02600; CAA26446.1; -  
 DR EMBL; X02600; CAA26447.1; -  
 DR EMBL; X13719; CAA31997.1; -  
 DR EMBL; X54147; CAA38086.1; -  
 DR PIR; A03324; OEPFPC.  
 DR MEROPS; A11.001; -  
 DR FlyBase; FBgn0013437; copia\GIP.  
 DR InterPro; IPR001969; Asprotease\_sile.  
 DR InterPro; IPR001584; Rye.  
 DR InterPro; IPR001878; Znf\_CCHC.  
 DR Pfam; PF00665; Iye; 1.  
 DR Pfam; PF00098; zf\_CCHC; 1.  
 DR PRINTS; PR00939; C2HCZNFINGER.  
 DR SMART; SM00343; ZNF\_C2HC; 1.  
 DR PROSITE; PS00141; ASP\_PROTEASE; FALSE\_NEG.  
 DR PROSITE; PS00158; ZF\_CCHC; 1.  
 DR TRANSPOSABLE element; Hydrolase; Aspartyl protease; ATP-binding;  
 KW Polypeptide; Alternative splicing; Polymorphism; Zinc-finger.  
 FT CHAIN 1 270  
 FT CHAIN 1 270  
 FT ZN FING 230 247  
 FT ACT\_SITE 292 292  
 FT VASPLIC 392 1374  
 FT FT  
 FT VARIANT 1265 1288  
 FT FT  
 FT VARIANT 1289 1409  
 FT MUTAGEN 292 292  
 FT CONFLICT 191 191  
 FT CONFLICT 300 300  
 FT CONFLICT 866 866  
 FT SEQUENCE 1409 AA; 162817 MW; BE89440763AA7691 CRC64;  
 Query Match 11.4%; Score 72; DB 1; Length 1409;  
 Best Local Similarity 23.5%; Pred. No. 13;  
 Matches 24; Conservative 18; Mismatches 34; Indels 26; Gaps 5;  
 QY 17 WFOED---NRCYIDNGKLFARSGIVGNMRFVDPKADYGVGENTLV--HADYVEVP 71  
 DB 1048 WFEFEQALKECEVY-----SSVRCIT--ILDKGNINENIVLLYVDVVIAT 1095  
 QY 72 GESLKNV-----VNLDPWPIFETLALRLVQGDVWL 104  
 DB 1096 GDMTRMNNFKRYLMKFRMTDLNFKPIGIRIMQEDKIYL 1137

RESULT 6  
 CDGT\_BACOH STANDARD; PRT; 704 AA.  
 AC P27036;  
 DT 01-NOV-1992 (Rel. 23, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 01-NOV-1995 (Rel. 32, Last annotation update)  
 DE Cyclomalodextrin glucanotransferase precursor (EC 2.4.1.19)  
 DE (Cyclodextrin-glycosyltransferase) (CGTase).  
 GS CGT.  
 OS Bacillus oshimensis.  
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
 OX NCBI\_TaxId=1481;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92000599; PubMed=1368710;  
 RA Sin K.A., Nakamura A., Kobayashi K., Masaki H., Uozumi T.;  
 RT "Cloning and sequencing of a cyclodextrin glucanotransferase gene  
 RT from Bacillus oshimensis and its expression in Escherichia coli."; Appl. Microbiol. Biotechnol. 35:600-605(1991).  
 CC -1- CATALYTIC ACTIVITY: Degrades starch to cyclodextrins by formation  
 CC of a 1,4-alpha-D-glucosidic bond.  
 CC -1- COFACTOR: BINDS TWO CALCIUM IONS.  
 CC -1- SUBUNIT: Monomer.  
 CC -1- MISCELLANEOUS: CGTASE MAY CONSIST OF TWO PROTEIN DOMAINS: THE ONE  
 CC IN THE AMINO-TERMINAL SIDE CLEAVES THE ALPHA-1,4-GLUCOSIDIC BOND  
 CC IN STARCH, AND THE OTHER IN THE C-TERMINAL SIDE CATALYZES OTHER  
 CC ACTIVITIES, INCLUDING THE RECONSTITUTION OF AN  
 CC ALPHA-1,4-GLUCOSIDIC LINKAGE FOR CYCLIZING THE  
 CC MALTOOLIGOSACCHARIDE PRODUCED.  
 CC -1- SIMILARITY: BELONGS TO FAMILY 13 OF GLYCOSYL HYDROLASES, ALSO  
 CC KNOWN AS THE ALPHA-AMYLASE FAMILY.

EMBL; D90243; BAA14289.2; -  
 DR HSSP; P31797; ICGF.  
 DR InterPro; IPR006589; Alp\_amy1\_cat\_sub.  
 DR InterPro; IPR006048; Alpha\_amy1\_C.  
 DR InterPro; IPR006047; Alpha\_amy1\_cat.  
 DR InterPro; IPR002044; CBD\_4.  
 DR InterPro; IPR006046; Glyco\_hydro\_13.  
 DR InterPro; IPR002909; IPT\_TIG.  
 DR Pfam; PF00128; alpha-amy1ase; 1.  
 DR Pfam; PF02806; alpha-amy1ase\_C; 1.  
 DR Pfam; PF00666; CBM\_20; 1.  
 DR Pfam; PF01833; TIG; 1.  
 DR PRINTS; PR00110; ALPHAAMYLASE.  
 DR PRODOM; PD001568; CBD\_4; 1.  
 DR SMART; SM00642; Amy; 1.  
 DR SMART; SM00632; Amy\_C; 1.  
 DR TRANSFERASE; Glycosyltransferase; Calcium; Signal.  
 FT SIGNAL 1 29  
 FT CHAIN 30 704  
 FT DISULFID 68 75  
 FT ACT\_SITE 251 251  
 FT ACT\_SITE 279 279  
 FT ACT\_SITE 350 350  
 FT SEQUENCE 704 AA; 78621 MW; 04FA14951D5ACB CB CRC64;  
 Query Match 11.2%; Score 70.5; DB 1; Length 704;  
 Best Local Similarity 25.0%; Pred. No. 8.7;  
 Matches 26; Conservative 21; Mismatches 44; Indels 13; Gaps 4;  
 QY 12 TAKHPFQIEDNRCYIDNGKLFARSGIVGNMRFVDPKADYGVGENTLVHADYVEVP 71



Db 159 TPHSSPALETDPSTYKENGAVYNDGVLIGNYN---Dp-----NNLFPHNGIDTPSS 207

Qy 72 GE-SLKCNVRNLDMPEIFETLRLRVLGQDVIML-RCVPELRVD 113

Dd 208 YEKSIYRNLVDLADYDINNTWMDQYIKBSIKMLMKGIDGIRVD 251

## RESULT 7

ID	US30_HCMVA	STANDARD;	PRT;	349 AA.
AC	P09706;			
DT	01-MAR-1989 (Rel. 10, Created)			
DT	01-MAR-1989 (Rel. 10, Last sequence update)			
DT	01-FEB-1991 (Rel. 17, Last annotation update)			
DE	Hypothetical protein HHRFS.			
GN	US30.			
OS	Human cytomegalovirus (strain AD169).			
OC	Viruses; dsDNA viruses, no RNA stage; Herpesviridae;			
OC	Betaherpesvirinae; Cytomegalovirus.			
OX	NCBI_taxid=10360;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=87169717; PubMed=3031311;			
RA	Weston K., Barrill B.G.;			
RT	"sequence of the short unique region, short repeats, and part of the			
RT	long repeats of human cytomegalovirus.";			
RL	J. Mol. Biol. 192:177-208(1986).			
RN	[2]			
RP	COMPLETE GENOME.			
RX	MEDLINE=90269039; PubMed=2161319;			
RA	Chee M.S., Bankier A.T., Beck S., Bohni R., Brown C.M., Cerny R.,			
RA	Horsnell T., Hutchinson C.A. III, Kouzarides T., Martignetti J.A.,			
RA	Preddie E., Satchwell S.C., Tomlinson P., Weston K.M., Barrill B.G.;			
RT	"analysis of the protein-coding content of the sequence of human			
RT	cytomegalovirus strain AD169."			
RL	Curr. Top. Microbiol. Immunol. 154:125-169(1990).			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (see <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>			
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).			

[illegible]

## RESULT 8

FX16_HUMAN	STANDARD;	PRT;	292 AA
ID	FX16_HUMAN		
CD	CD163		

[illegible]

## RESULT 9

ID	GAL7_CRYNE	STANDARD;	PRT;	381 AA.
AC	P40908;			
DT	01-FEB-1995 (Rel. 31, Created)			
DT	01-FEB-1995 (Rel. 31, Last sequence update)			
DT	15-SEP-2003 (Rel. 42, Last annotation update)			
DE	Galactose-1-phosphate uridylyltransferase (EC 2.7.7.12) (Gal-1-P			
DE	uridylyltransferase) (UDP-glucose--hexose-1-phosphate			
DE	uridylyltransferase).			
GN	GAL7.			
OS	Cryptococcus neoformans (Filobasidiella neoformans).			
OC	Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Heterobasidiomycetes;			
OC	Tremellalomycetidae; Tremellales; Tremellaceae; Filobasidiella.			
OX	NCBI_TaxID=5507;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=96020643; PubMed=8577246;			
RA	Wickes B.L., Edman J.C.;			
RT	"The Cryptococcus neoformans GAL7 gene and its use as an inducible			
RT	promoter.";			
RL	Mol. Microbiol. 16:1099-1109(1995).			
CC	-1- CATALYTIC ACTIVITY: UDP-glucose + alpha-D-galactose 1-phosphate =			
CC	alpha-D-glucose 1-phosphate + UDP-galactose.			

```
CC -1- COFACTOR: Binds 1 zinc and 1 iron ion per subunit (By similarity).
CC -1- PATHWAY: Galactose metabolism; second step.
CC -1- SIMILARITY: Belongs to the galactose-1-phosphate
CC      uridylyltransferase family 1.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U16994; AAA92683.1; -.
CC PIR: S69795; S69795.
CC HSSP: P09148; IHXP.
CC InterPro: IPR001937; GalP_UDPtransf1.
CC InterPro: IPR005851; GalP_Utranafe_1.
CC InterPro: IPR005850; GalP_Utranafe_C.
CC InterPro: IPR005849; GalP_Utranafe_N.
CC Pfam: PF02744; GalP_UDP_tf_C_1.
CC Pfam: PF01087; GalP_UDP_transf_1.
CC Prodom: PD005051; GalP_UDPtransf1; 1.
CC TIGRFAMs: TIGR00209; galT_1; 1.
CC PROSITE: PS00117; GAL_P_UDP_TRANSF_1; 1.
CC KW Transferrase; Nucleotidyltransferase; Galactose metabolism; Zinc; Iron;
KW Metal-binding.
CC ACT_SITE 188 188 NUCLEOPHILE (BY SIMILARITY).
CC FT METAL 65 65 ZINC (BY SIMILARITY).
CC FT METAL 68 68 ZINC (BY SIMILARITY).
CC FT METAL 131 131 ZINC (BY SIMILARITY).
CC FT METAL 186 186 ZINC (BY SIMILARITY).
CC FT METAL 204 204 IRON (BY SIMILARITY).
CC FT METAL 306 306 IRON (BY SIMILARITY).
CC FT METAL 323 323 IRON (BY SIMILARITY).
CC FT METAL 325 325 IRON (BY SIMILARITY).
CC SEQUENCE 381 AA; 43032 MW; 49119349C2A4463 CRC64;
SO
Query Match 10.7%; Score 67; DB 1; Length 381;
Best Local Similarity 22.6%; Pred. No. 10;
Matches 21; Conservative 15; Mismatches 39; Indels 18; Gaps 3;
QY 5 ALRPWKSTAHKPMFQIE---DNRCYINGKLFARGSIIVGNMSRFVDPKADYGVGCVENLY 61
DB 42 SLRPWNQKSTPAIPVTPPHDSKCYLCRGN-----KRTTGHANDYKGI-----Y 86
QY 62 VHADVEFVPGESLKNVRLDVMPIFETLALR 94
DB 87 VFENDFPALLPDPLAVGTNKSIDPLFQSEFVR 119
RESULT 10
AD08 HUMAN STANDARD; PRT; 824 AA.
ID AD08 HUMAN STANDARD; PRT; 824 AA.
AC P78325;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE ADAM 8 precursor (EC 3.4.24.-) (A disintegrin and metalloprotease
DE domain 8) (Cell surface antigen MS2) (CD156a antigen) (CD156).
DE ADAM8 OR MS2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RC SEQUENCE FROM N.A.
RC TISSUE=Blood.
RX MEDLINE=97271556; PubMed=9126482;
RA Yoshijama K., Higuchi Y., Kataoka M., Matsuura K., Yamamoto S.;
RT "CD156 (human ADAM8): expression, primary amino acid sequence, and
RT gene location.";
RL Genomics 41:56-62(1997).
```

```
CC -1- FUNCTION: POSSIBLE INVOLVEMENT IN EXTRAVASATION OF LEUKOCYTES.
CC -1- COFACTOR: Binds 1 zinc ion per subunit (probable).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- TISSUE SPECIFICITY: EXPRESSED ON NEUTROPHILS AND MONOCYTES.
CC -1- SIMILARITY: Belongs to peptidase family M12B.
CC -1- SIMILARITY: Contains 1 disintegrin domain.
CC -1- DATABASE: NAME=PROW; NOTE=CD guide CD156 entry;
CC      WWW:http://www.ncbi.nlm.nih.gov/prow/cd/cd156.htm.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: D6579; BAA05626.1; -.
CC HSSP: P18619; IFVL.
CC MEROPS: M12.208; -.
CC Genew: HGNC:215; ADAM8.
CC MIM: 602267; -.
CC GO: GO:0005888; C:cytoecglycan integral to plasma membrane; TAS.
CC InterPro: IPR006386; ADAM_cysteine.
CC InterPro: IPR001762; Disintegrin.
CC InterPro: IPR006209; EGF-like.
CC InterPro: IPR006210; IEGF.
CC InterPro: IPR002870; Pep_M12B_propep.
CC InterPro: IPR001590; Repolysin.
CC InterPro: IPR006025; Zn_MTPeptidase.
CC Pfam: PF00200; disintegrin; 1.
CC Pfam: PF01562; Pep_M12B_propep; 1.
CC Pfam: PF01421; Repolysin; 1.
CC PRINTS: PR00289; DISINTEGRIN.
CC PRODOM: PD000664; Disintegrin; 1.
CC SMART: SM00608; ACR; 1.
CC SMART: SM00050; DISIN; 1.
CC PROSITE: PS50215; ADAM_MEROP; 1.
CC PROSITE: PS00427; DISINTEGRIN_1; FALSE_NEG.
CC PROSITE: PS50214; DISINTEGRIN_2; 1.
CC PROSITE: PS00142; ZINC_PROTEASE; 1.
CC PROSITE: PS00142; ZINC_PROTEASE; 1.
CC KW Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein;
KW Transmembrane; Antigen.
CC FT SIGNAL 1 16 POTENTIAL.
CC FT CHAIN 17 824 ADAM 8.
CC FT DOMAIN 17 685 EXTRACELLULAR (POTENTIAL).
CC FT TRANSMEM 656 676 POTENTIAL.
CC FT DOMAIN 677 824 CYTOPLASMIC (POTENTIAL).
CC FT DOMAIN 200 400 METALLOPROTEASE.
CC FT DOMAIN 408 494 DISINTEGRIN-LIKE.
CC FT METAL 334 334 ZINC (CATALYTIC) (PROBABLE).
CC FT ACT_SITE 335 335 BY SIMILARITY.
CC FT METAL 338 338 ZINC (CATALYTIC) (PROBABLE).
CC FT METAL 344 344 ZINC (CATALYTIC) (PROBABLE).
CC FT METAL 344 344 ZINC (CATALYTIC) (PROBABLE).
CC FT DISUFID 310 395 BY SIMILARITY.
CC FT CARBOHYD 67 67 N-LINKED (GLCNAC...) (POTENTIAL).
CC FT CARBOHYD 91 91 N-LINKED (GLCNAC...) (POTENTIAL).
CC FT CARBOHYD 436 436 N-LINKED (GLCNAC...) (POTENTIAL).
CC FT CARBOHYD 612 612 N-LINKED (GLCNAC...) (POTENTIAL).
CC SEQUENCE 824 AA; 88673 MW; 5DF8E05F30DF479E CRC64;
SO
Query Match 10.7%; Score 67; DB 1; Length 824;
Best Local Similarity 26.4%; Pred. No. 25;
Matches 29; Conservative 18; Mismatches 35; Indels 28; Gaps 6;
QY 14 KHPWFOIED-----NRCYIDNGKLFARGSIIVGNMSRFVDPKADYGVGCVEN----- 59
DB 151 RHAIVYQAEHLQAGNGCGVSDSL---GSLIGRTAAVRRPR-----GDSLPSRRTRYV 202
QY 60 -LYVHADVEF-VPGESLKNVRLDVM---PIFETLALRLVLQGDVIT 103
```

DB		203 ELYVVVDNAEFQMLGSEAAVSHRVLEYVNHDVKLYQKLNFRVLVGLIELW	252
RESULT 11			
ID	UHPT_SALTY	STANDARD;	PRT; 463 AA.
AC	P27670:		
DT	01-AUG-1992 (Rel. 23, Created)		
DT	28-FEB-2003 (Rel. 41, Last sequence update)		
DT	28-FEB-2003 (Rel. 41, Last annotation update)		
DE	Hexose phosphate transport protein.		
GN	UHPT OR STM3787.		
OS	Salmonella typhimurium.		
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;		
OC	Enterobacteriaceae; Salmomella.		
OX	NCBI_TaxID=602;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=Lt2;		
RC	MEDLINE=92234930; PubMed=1569007;		
RA	Island M.D., Wei B.-Y., Kacher R.J.;		
RT	"Structure and function of the uhp genes for the sugar phosphate		
RL	transport system in Escherichia coli and Salmonella typhimurium.";		
RN	J. Bacteriol. 174:2754-2762(1992).		
RP	[2]		
RC	SEQUENCE FROM N.A.		
RC	STRAIN=Lt2 / SGSC1412 / ATCC 700720;		
RA	MEDLINE=21534948; PubMed=11677609;		
RA	McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,		
RA	Courtney L., Porwollik S., Ali U., Dante M., Du F., Hou S., Layman D.,		
RA	Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,		
RA	Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,		
RA	Waterston R., Wilson R.K.;		
RT	"Complete genome sequence of Salmonella enterica serovar Typhimurium		
RL	Lt2.";		
RP	Nature 413:852-856(2001).		
CC	-I- FUNCTION: TRANSPORT PROTEIN FOR SUGAR PHOSPHATE UPTAKE.		
CC	-I- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane.		
CC	-I- INDUCTION: EXTERNAL GLUCOSE-6-PHOSPHATE INDUCES THE EXPRESSION		
CC	OF THE UHP-REGION.		
CC	-I- SIMILARITY: BELONGS TO THE SLC37A FAMILY OF TRANSPORTERS.		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (see http://www.isb-sib.ch/announce/		
CC	or send an email to license@isb-sib.ch).		
CC	-----		
DR	EMBL; M89480; AAA27246.1; --		
DR	EMBL; AE008876; AAL22645.1; --		
DR	PIR; D41853; D41853.		
DR	StyGene; SG10409; ubpt.		
DR	InterPro; IPR000849; GlpT_transporter.		
DR	InterPro; IPR007114; MFS.		
DR	InterPro; IPR005828; Sub_transporter.		
DR	Pfam; PF00083; sugar_tr_1.		
DR	TIGRFAMs; TIGR00881; 2A0104; 1.		
DR	PROSITE; PS00942; GLPT; 1.		
KM	Transport; Sugar transport; Transmembrane; Inner membrane;		
KM	Complete proteome.		
FT	DOMAIN 1 26 CYTOPLASMIC (PROBABLE) .		
FT	TRANSSEM 27 45 PROBABLE .		
FT	DOMAIN 46 66 PERIPLASMIC (PROBABLE) .		
FT	TRANSSEM 67 89 PROBABLE .		
FT	DOMAIN 90 96 CYTOPLASMIC (PROBABLE) .		
FT	TRANSSEM 97 116 PROBABLE .		
FT	DOMAIN 117 120 PERIPLASMIC (PROBABLE) .		
FT	TRANSSEM 121 140 PROBABLE .		
FT	DOMAIN 141 159 CYTOPLASMIC (PROBABLE) .		
FT	TRANSSEM 160 187 PROBABLE .		
FT	DOMAIN 188 190 PERIPLASMIC (PROBABLE) .		

```

FT TRANSMEM 191 210 PROBABLE. CYTOPLASMIC (PROBABLE).
FT DOMAIN 211 249
FT TRANSMEM 250 274
FT TRANSMEM 250 294 PERIPLASMIC (PROBABLE).
FT TRANSMEM 295 318 PERIPLASMIC (PROBABLE).
FT TRANSMEM 319 327 CYTOPLASMIC (PROBABLE).
FT TRANSMEM 328 344 PERIPLASMIC (PROBABLE).
FT DOMAIN 345 350 PERIPLASMIC (PROBABLE).
FT TRANSMEM 351 379 PROBABLE.
FT DOMAIN 380 391 CYTOPLASMIC (PROBABLE).
FT TRANSMEM 392 411 PERIPLASMIC (PROBABLE).
FT TRANSMEM 412 425 PERIPLASMIC (PROBABLE).
FT TRANSMEM 426 446 PROBABLE.
FT DOMAIN 447 463 CYTOPLASMIC (PROBABLE).
FT CONFLICT 16 16 P -> A (IN REF. 1).
FT CONFLICT 385 386 GA -> AL (IN REF. 1).
FT CONFLICT 413 413 G -> A (IN REF. 1).
SQ SEQUENCE 463 AA; 50708 MM; C13398B21CEA92DA CRC64;

Query Match 10.6%; Score 66.5; DB 1; Length 463;
Best Local Similarity 31.1%; Pred. No. 14;
Matches 23; Conservative 7; Mismatches 27; Indels 17; Gaps 4;

QY 38 IGVNMS-RFVPPDPADYG-GVGENLY---VHADDFEVPQSGSLKWNVRNLDMVWIFETLA 92
Db 204 IGVGTGLRFGSDSPESYGLKAEELFEELISEEDKTEVENMTKMQI-----F 251
QY 93 LRLVLQGDVIMLRC 106
Db 252 VEYVLKKNVIMLRC 265

RESULT 12
ATC7 YEAST STANDARD; PRT; 1151 AA.
AC P40527;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Potential phospholipid-transporting ATPase 4 (EC 3.6.3.1).
GN NEO1 OR YIL048W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RX PubMed=9169870;
RA Churcher C.M., Bowman S., Badcock K., Bankier A., Brown D.,
RA Chillingworth T., Connor R., Devlin K., Gentles S., Hamlin N.,
RA Harris D.E., Horsnell T., Hunt S., Jagsels K., Jones M., Lye G.,
RA Moulé S., Odell C., Pearson D., Rajandream M.A., Rice P., Rowley N.,
RA Skelton J., Smith V., Walsh S., Whitehead S., Barrrell B.G.;
RT "The nucleotide sequence of Saccharomyces cerevisiae chromosome IX.";
RL Nature 387:84-87 (1997).
CC -1- FUNCTION. THIS MAGNESIUM DEPENDENT ENZYME CATALYZES THE HYDROLYSIS
CC OF ATP COUPLED WITH THE TRANSPORT OF PHOSPHOLIPIDS (POTENTIAL).
CC LEADS TO NEOMYCIN-RESISTANCE WHEN OVEREXPRESSED.
CC -1- CATALYTIC ACTIVITY: ATP + H(2)O = ADP + phosphate.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- SIMILARITY: Belongs to the cation transport ATPases family (P-type
CC ATPases). Subfamily IV.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announcements
CC or send an email to license@ebi.ac.uk).
CC -----
CC EMBL; Z38060; CAA86174.1; -.

```

```

DR PIR; S48431; S48431.
DR SGD; S0001310; NBD1.
DR InterPro; IPR001577; ATPase_E1-E2.
DR InterPro; IPR006539; Flippase.
DR InterPro; IPR005834; Hydrolase.
DR Pfam; PF00122; E1-E2_ATPase; 1.
DR Pfam; PF00702; Hydrolase; 1.
DR PRINTS; PR00119; CATNPASE.
DR TIGRFA; TIGR01652; ATPase-plipid; 1.
DR TIGRFA; TIGR01494; ATPase_P-type; 7.
DR PROSITE; PS00154; ATPASE_E1_E2; 1.
KW Hydrolase; Transmembrane; Phosphorylation; Magnesium; ATP-binding.
SQ
SEQUENCE 1151 AA; 130217 MW; DC7225CC9577DBB6 CRC64;

Query Match
Best Local Similarity 10.6%; Score 66.5; DB 1; Length 1151;
Matches 26; Conservative 24; Mismatches 47; Indels 11; Gaps 5;

QY 6 LRPKSTAHKHPQIDNDRCYINDGKLFARGSVGNMRFVDPKADYGVGENTLVHAD 65
DB 768 IKTWMLTGD---KVFARVGSISAKLISRGQYVHTTK-VTRPBGAFNOL-FYLKINRN 821
QY 66 DVEFVGEISLKMVNR-NLDMPIFETLA RLVLQGDVIMLRCPVELRVD 113
DB 822 ACCLIGESLGMFLKHYE-QEFPDV---VHLPTYLACKCTPQQRAD 864

RESULT 13
KAD STRMU STANDARD; PRT; 212 AA.
AC QBD533;
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DB Adenylate kinase (EC 2.7.4.3) (ATP-AMP transphosphorylase).
GN ADK OR SMU 2005.
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OC NCBI_TaxID=1309;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-UN159 / ATCC 700610 / Serotype C;
RA MEDLINE=2235063; PubMed=12397196;
RA Adic D., Moshan W.M., McLaughlin R.E., Savic G., Chang J.,
RA Carson M.B., Primeaux C., Tian R., Kenton S., Jia H., Lin S., Qian Y.,
RA Li S., Zhu H., Najaf P., Lai H., White J., Roe B.A., Ferretti J.J.;
RT "Genome sequence of Streptococcus mutans UN159, a cariogenic dental
pathogen."
RL Proc. Natl. Acad. Sci. U.S.A. 99:14434-14439(2002).
CC -1- FUNCTION: This small ubiquitous enzyme is essential for
CC maintenance and cell growth.
CC -1- CATALYTIC ACTIVITY: ATP + AMP = ADP + ADP.
CC -1- SUBUNIT: Monomer (By similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- SIMILARITY: Belongs to the adenylate kinase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

```

```

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE015024; AAN59609.1; -.
DR HAMAP; MF_00235; -; 1.
DR Pfam; PF00406; ADK; 1.
DR ProDom; PD000657; Adenylate kin; 1.
DR TIGRFA; TIGR01351; adk; 1.
DR PROSITE; PS00113; ADENYLATE_KINASE; 1.
KW Transferrase; Kinase; ATP-binding; Complete proteome.
FT NP BIND 7 15
SQ SEQUENCE 212 AA; 23638 MW; 5AF9C1AD9B5D5520 CRC64;

Query Match
Best Local Similarity 10.5%; Score 66; DB 1; Length 212;
Matches 22; Conservative 18; Mismatches 27; Indels 26; Gaps 5;

QY 22 DNRCTYID--NGKLFARGSVGNMRFVDPKADYGVGENTLVHAD----- 66
DB 117 DPACLVRLSGRIINRKT--GETYHKVFNPADYV--EDDYQREDDKPEYVKRLDVNI 172
QY 67 -----VEFVGEISLKMVNR-NLDMPIFETLA 92
DB 173 AQGHPIIEYRNKGLVYDIEGNQDINLVFETIA 205

RESULT 14
GLNA PANAR STANDARD; PRT; 361 AA.
AC 004831;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Glutamine synthetase (EC 6.3.1.2) (Glutamate--ammonia ligase).
OS Panulirus argus (Spiny lobster).
OC Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;
OC Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Palinura;
OC Palinuroidea; Palinuridae; Panulirus.
OC NCBI_TaxID=6737;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Olfactory organ;
RA MEDLINE=93314973; PubMed=8100791;
RA Trapido-Rosenchal H.G., Liner P.J., Greenberg R.M., Gleason R.A.,
RA Carr W.E.;
RT "cDNA clones from the olfactory organ of the spiny lobster encode a
RT protein related to eukaryotic glutamine synthetase."
RL Gene 129:275-278(1993).
CC -1- CATALYTIC ACTIVITY: ATP + L-glutamate + NH(3) = ADP + phosphate +
CC L-glutamine.
CC -1- SUBUNIT: Homooctamer.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: BELONGS TO THE GLUTAMINE SYNTHETASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M6796; AAA02583.1; -.
DR PIR; JN0716; JN0716.
DR HSSP; P06201; ILGR.
DR InterPro; IPR001691; GLN_synth.
DR InterPro; IPR001637; GLN_adenyltn.
DR Pfam; PF03920; gln-synt; 1.
DR Pfam; PF03951; gln-synt_N; 1.
DR ProDom; PD001057; Gln_synt_C; 1.
DR PROSITE; PS00180; GLNA_1; 1.
DR PROSITE; PS00181; GLNA_ATP; 1.

```







```
XX 02-FEB-2000; 2000MO-US02740.
PF
XX 02-FEB-1999; 99US-0118287.
XX (UYOR-) UNIV OREGON HEALTH SCI.
XX Johnson DC, Tomazin R, Boname J, Hegde NR;
PI WPI; 2000-506069/45.
DR N-PSDB; AAA53814.
XX
XX Inhibiting recognition of cellular tissue by CD8+ and CD4+ T cells, to
PT treat or prevent autoimmune diseases, and to improve gene therapy,
PT comprises introducing human cytomegalovirus US2 protein into cells
XX
XX Example 12; Page 49; 53pp; English.
XX
XX Human cytomegalovirus (hCMV) US2 protein, which has previously been shown
CC to block the major histocompatibility complex (MHC) class I antigen
CC presentation pathway, blocks the MHC class II pathway. The US2 protein
CC does not have to be mutated to cause inhibition of the MHC class II
CC pathway. The binding domain recognizes MHC I heavy chains, MHC II alpha
CC chains, and optionally DM-alpha chains. US2 has a double inhibitory
CC effect on the MHC class II pathway, inhibiting recognition of cellular
CC tissue by CD8-positive and CD4-positive T cells. US2 or its soluble
CC variants, can be used to reduce inappropriate immune responses. The US2
CC protein can be used to improve the persistence of a virus. Vectors
CC encoding soluble US2 protein (residues 28-143) can be used to treat
CC autoimmune disease, especially where it is mediated by MHC II molecules.
CC The vector can also be used to improve gene therapy, and preferably also
CC contains a sequence encoding a therapeutic product. The protein is
CC exogenously supplied or expressed from a recombinant cell, and
CC may also be used to inhibit CD4-positive mediated immune responses,
CC autoimmune responses, transplant immune responses and gene therapy
CC immune responses.
XX
XX Sequence 177 AA;
SQ
Query Match 100.0%; Score 947; DB 21; Length 177;
Best Local Similarity 100.0%; Pred. No. 7.9e-99;
Matches 177; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MKFLVNAVLFVWVYISYIYARLPDGIITKAGEDALRPWKSTAKHPWFOIEDNRCYINGK 60
DB 1 MKFLVNAVLFVWVYISYIYARLPDGIITKAGEDALRPWKSTAKHPWFOIEDNRCYINGK 60
QY 61 LFARGSIVGNMSRFVDPKADYGVGENLYVHADVEFVGESLKNVNRNLDPVPIFETL 120
DB 61 LFARGSIVGNMSRFVDPKADYGVGENLYVHADVEFVGESLKNVNRNLDPVPIFETL 120
QY 121 ALRLVLOGDYIWLRCVPELRYDVTSSAYMMNMQYGMVGOPELAPEDDESDALLDPV 177
DB 121 ALRLVLOGDYIWLRCVPELRYDVTSSAYMMNMQYGMVGOPELAPEDDESDALLDPV 177
RESULT 2
AAI97248
ID AAY97248 standard; Protein; 199 AA.
XX
XX AAY97248;
XX
XX 04-DEC-2000 (first entry)
XX
XX Cytomegalovirus US2 protein.
XX
XX US2; hCMV; major histocompatibility complex; MHC; class I; class II;
XX antigen presentation; inhibition; CD8-positive; CD4-positive; T cell;
XX transplant; gene therapy; immunosuppressive.
XX
XX Human cytomegalovirus.
XX
XX WO200046361-A1.
```

```
XX 10-AUG-2000.
PD
XX
XX 02-FEB-2000; 2000MO-US02740.
PF
XX 02-FEB-1999; 99US-0118287.
XX (UYOR-) UNIV OREGON HEALTH SCI.
XX Johnson DC, Tomazin R, Boname J, Hegde NR;
PI WPI; 2000-506069/45.
DR N-PSDB; AAA53813.
XX
XX Inhibiting recognition of cellular tissue by CD8+ and CD4+ T cells, to
PT treat or prevent autoimmune diseases, and to improve gene therapy,
PT comprises introducing human cytomegalovirus US2 protein into cells
XX
XX Claim 6; Page 47-48; 53pp; English.
XX
XX This is the human cytomegalovirus (hCMV) US2 protein, which has
CC previously been shown to block the major histocompatibility complex (MHC)
CC class I antigen presentation pathway, blocks the MHC class II pathway.
CC The US2 protein does not have to be mutated to cause inhibition of the
CC MHC class II pathway. The binding domain recognizes MHC I heavy chains,
CC MHC II alpha chains, and optionally DM-alpha chains. US2 has a double
CC inhibitory effect on the MHC class II pathway, inhibiting recognition of
CC cellular tissue by CD8-positive and CD4-positive T cells. US2 or its
CC soluble variants, can be used to reduce inappropriate immune responses.
CC The US2 protein can be used to improve the persistence of a virus.
CC Vectors encoding soluble US2 protein (residues 28-143) can be used to
CC treat autoimmune disease, especially where it is mediated by MHC II
CC molecules. The vector can also be used to improve gene therapy, and
CC preferably also contains a sequence encoding a therapeutic product. The
CC protein is exogenously supplied or expressed from a recombinant cell, and
CC may also be used to inhibit CD4-positive mediated immune responses,
CC autoimmune responses, transplant immune responses and gene therapy
CC immune responses.
XX
XX Sequence 199 AA;
SQ
Query Match 78.1%; Score 740; DB 21; Length 199;
Best Local Similarity 100.0%; Pred. No. 2.5e-75;
Matches 196; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 22 RLDPGITKAGEDALRPWKSTAKHPWFOIEDNRCYINGKLFARGSIVGNMSRFVDPKAD 81
DB 20 RLDPGITKAGEDALRPWKSTAKHPWFOIEDNRCYINGKLFARGSIVGNMSRFVDPKAD 79
QY 82 YGVGENLYVHADVEFVGESLKNVNRNLDPVPIFETLALRLVLOGDYIWLRCVPELRY 141
DB 80 YGVGENLYVHADVEFVGESLKNVNRNLDPVPIFETLALRLVLOGDYIWLRCVPELRY 139
QY 142 DYTSSAYMMNMQYGMV 157
DB 140 DYTSSAYMMNMQYGMV 155
RESULT 3
ABG31298
ID ABG31298 standard; Protein; 451 AA.
XX
XX ABG31298;
XX
XX 21-OCT-2002 (first entry)
XX
XX Hydrophobic protein melittin tag tag-human beta2-adrenergic receptor EF.
XX
XX Signal sequence; ligand identification; hydrophobic protein;
XX transmembrane protein; monocytic membrane protein;
XX polycyclic membrane protein; pump; channel; receptor kinase;
XX G protein-coupled receptor; transporter protein.
XX
```



OS Synthetic.  
 XX PN MO200257792-A2.  
 XX PD 25-JUL-2002.  
 XX PF 19-DEC-2001; 2001MO-US50088.  
 XX PR 29-DEC-2000; 2000US-258970P.  
 XX PA (NEOG-) NEOGENESIS PHARM INC.  
 XX PI Felsch JS, Annis DA, Kalghatgi K, Nash HM;  
 XX DR WPI; 2002-599728/64.  
 XX PT Identifying ligand for hydrophobic protein based on affinity selection  
 XX PT which can operate in the presence of amphiphile without regard to the  
 XX PT specific biological function of hydrophobic target protein -  
 XX PS Disclosure; Fig 2; 97pp; English.

CC This invention relates to a novel method for identifying a ligand for a  
 CC hydrophobic protein. The method comprises selecting a ligand molecule by  
 CC affinity selection by exposing a hydrophobic target protein bound by an  
 CC amphiphile to a multiplicity of molecules to promote formation of at  
 CC least a complex between the hydrophobic target protein and the ligand  
 CC molecule, separating the complex from the unbound molecules, and  
 CC identifying the ligand molecule. The method of the invention is  
 CC useful for identifying a ligand for hydrophobic protein such as a  
 CC membrane, integral membrane, transmembrane, monotopic or polytopic  
 CC or transporter protein, or membrane-associated enzyme, or Myc tag-EB  
 CC tag-human m2 mChR, flag tag-human beta2 adrenergic receptor-EB tag,  
 CC human neurokinin 3 receptor-HSV tag-Myc tag, flag tag-human m1 mChR-EB  
 CC tag, and rat m3 mChR-HSV tag-Octahis tag. The ligand identified by the  
 CC method of the invention is useful for the development of novel  
 CC medicines and medicinal diagnostics. The present sequence represents  
 CC the hydrophobic protein melittin tag tag-human beta2-adrenergic receptor  
 CC EB used in the method of the invention.

CC Sequence 451 AA;  
 XX SQ

Query Match 12.4%; Score 117.5; DB 23; Length 451;  
 Best Local Similarity 27.7%; Pred. No. 0.00022;  
 Matches 41; Conservative 13; Mismatches 45; Indels 49; Gaps 3;

QY 1 MKELVNVAVLVFVWVYISYIYARLPDGITKAGEDALRPWKSTAKHPWQIEDNRCYIDNGK 60  
 DB 1 MKELVNVAVLVFVWVYISYIYARLPDGITKAGEDALRPWKSTAKHPWQIEDNRCYIDNGK 60  
 QY 61 LPARGSIVGNMSPFVDDPKADYGVGNGENLYHADYVEFVPGESLKNVNRNLDVMPFETL 120  
 DB 34 -----GNGSAFLAPNRS-----HAPDHVTCQORDEVVWG---MGIVMSL 71  
 QY 121 ALRLVLOGDVIMRCVPELRVDYTSSAY 148  
 DB 72 IYLAIVFANVLVITAIKFERLQTVTVY 99

RESULT 4  
 ABG31300 standard; Protein; 498 AA.  
 AC ABG31300;  
 DT 21-OCT-2002 (first entry)  
 XX DE Hydrophobic protein melittin-flag @tag-human m1 mChR-EB.  
 XX KW Signal sequence; ligand identification; hydrophobic protein;  
 XX KW transmembrane protein; monotopic membrane protein;  
 XX KW polytopic membrane protein; pump; channel; receptor kinase;

KW G protein-coupled receptor; transporter protein.  
 XX OS Synthetic.  
 XX PN MO200257792-A2.  
 XX PD 25-JUL-2002.  
 XX PF 19-DEC-2001; 2001MO-US50088.  
 XX PR 29-DEC-2000; 2000US-258970P.  
 XX PA (NEOG-) NEOGENESIS PHARM INC.  
 XX PI Felsch JS, Annis DA, Kalghatgi K, Nash HM;  
 XX DR WPI; 2002-599728/64.  
 XX PT Identifying ligand for hydrophobic protein based on affinity selection  
 XX PT which can operate in the presence of amphiphile without regard to the  
 XX PT specific biological function of hydrophobic target protein -  
 XX PS Disclosure; Fig 4; 97pp; English.

CC This invention relates to a novel method for identifying a ligand for a  
 CC hydrophobic protein. The method comprises selecting a ligand molecule by  
 CC affinity selection by exposing a hydrophobic target protein bound by an  
 CC amphiphile to a multiplicity of molecules to promote formation of at  
 CC least a complex between the hydrophobic target protein and the ligand  
 CC molecule, separating the complex from the unbound molecules, and  
 CC identifying the ligand molecule. The method of the invention is  
 CC useful for identifying a ligand for hydrophobic protein such as a  
 CC membrane, integral membrane, transmembrane, monotopic or polytopic  
 CC or transporter protein, or membrane-associated enzyme, or Myc tag-EB  
 CC tag-human m2 mChR, flag tag-human beta2 adrenergic receptor-EB tag,  
 CC human neurokinin 3 receptor-HSV tag-Myc tag, flag tag-human m1 mChR-EB  
 CC tag, and rat m3 mChR-HSV tag-Octahis tag. The ligand identified by the  
 CC method of the invention is useful for the development of novel  
 CC medicines and medicinal diagnostics. The present sequence represents  
 CC the hydrophobic protein melittin-flag @tag-human m1 mChR-EB used in the  
 CC method of the invention.

CC Sequence 498 AA;  
 XX SQ

Query Match 11.6%; Score 110; DB 23; Length 498;  
 Best Local Similarity 51.8%; Pred. No. 0.0018;  
 Matches 29; Conservative 0; Mismatches 13; Indels 14; Gaps 2;

QY 1 MKELVNVAVLVFVWVYISYIYARLPDGITKAGEDA-----LRPWSTAKHPW 46  
 DB 1 MKELVNVAVLVFVWVYISYIYARLPDGITKAGEDA-----LRPWSTAKHPW 46  
 QY 121 ALRLVLOGDVIMRCVPELRVDYTSSAY 148  
 DB 72 IYLAIVFANVLVITAIKFERLQTVTVY 99

RESULT 5  
 AAM47008 standard; Protein; 1189 AA.  
 AC AAM47008;  
 DT 13-AUG-1998 (first entry)  
 XX DE Glutathione-S-transferase and hTRT fusion protein 8.  
 XX KW Human; telomerase reverse transcriptase; hTRT; TRT; diagnosis;  
 XX KW prognosis; cell proliferation; cancer; ageing; ribonucleoprotein.  
 XX OS Synthetic.  
 XX OS Homo sapiens.  
 XX FT Key Location/Qualifiers  
 FT Misc-difference 22..23  
 FT /note= "enterokinase cleavage site"

XX GB2317891-A.  
 XX  
 XX 08-APR-1998.  
 XX  
 XX 01-OCT-1997; 97GB-0020890.  
 XX  
 XX 14-AUG-1997; 97US-0915503.  
 XX 01-OCT-1996; 96US-0724643.  
 XX 18-APR-1997; 97US-0844419.  
 XX 25-APR-1997; 97US-0846017.  
 XX 06-MAY-1997; 97US-0851843.  
 XX 09-MAY-1997; 97US-0854050.  
 XX 14-AUG-1997; 97US-0911312.  
 XX 14-AUG-1997; 97US-0912951.  
 XX  
 XX (GERO-) GERON CORP.  
 XX (UYTE-) UNIV TECHNOLOGY CORP.  
 XX  
 XX Andrews WH, Cech TR, Chapman KB, Harley C, Lingner J;  
 XX Morin GB, Nakamura T, Harley CB;  
 XX  
 XX WPI; 1998-171633/16.  
 XX  
 XX Pure and recombinant human Telomerase Reverse Transcriptase and its  
 XX variants - are useful in the diagnosis, prognosis and treatment of  
 XX cell proliferation conditions especially cancer and ageing  
 XX  
 XX Example 6; Page 234-235; 387pp; English.  
 XX  
 XX The present sequence represents a fusion protein from an example  
 XX of the present invention which describes human telomerase reverse  
 XX transcriptase (hTERT). The present invention also describes the  
 XX following methods: (A) determining whether a test compound is  
 XX a modulator of hTERT, by detecting the change in hTERT recombinant  
 XX protein or polynucleotide, on administration of the compound;  
 XX (B) preparation of recombinant telomerase by contacting a protein  
 XX preparation of hTERT with a telomerase RNA component; (C) detection  
 XX of the hTERT RNA or protein in a sample by binding a relevant  
 XX probe to the sample and detecting the complex formed or in the case of  
 XX RNA detection, amplifying the product and correlating the presence of  
 XX complex or amplification product with presence of hTERT in the sample;  
 XX and (D) increasing the proliferation of a vertebrate cell by increasing  
 XX hTERT expression; and (E) the use of an agent that causes an increase in  
 XX cell vertebrate cell proliferation to create a medicament that inhibits  
 XX ageing. A protein preparation of hTERT and the polynucleotide encoding  
 XX hTERT can be used in the manufacture of medicaments for inhibiting the  
 XX effect of ageing or cancer. Inhibitors of telomerase activity can be  
 XX used to treat conditions that are associated with high telomerase  
 XX activity. A protein preparation of hTERT can also be used in the new  
 XX methods.  
 XX  
 XX Sequence 1189 AA;  
 XX  
 QY Query Match 11.6%; Score 109.5; DB 19; Length 1189;  
 DB Best Local Similarity 32.1%; Pred. No. 0.0067;  
 Matches 50; Conservative 10; Mismatches 65; Indels 31; Gaps 6;  
 QY 1 MKFLVNVAVLVMVYISYIYARLPDGTGAG-EDA-----LIPKSTAKHPFQI 49  
 DB 1 MKFLVNVAVLVMVYISYIYAD-PSRSAGTMEFAASTQRCVLTMTWELAPATPAMP 59  
 QY 50 EDNRCYIDNGCLPARGSIGVGMRSFVDDPKADYGVGENLYHADDFE-----V 99  
 DB 60 RAPRCRAVRSLLRSHYREVPPLATFV-----RRLPQGMRLVQRGDPAPAFALVAQCLVCV 115  
 QY 100 PGESLKMNVRLDVMPFETLATRLVQGVYIMRC 135  
 DB 116 P-----WDARPPPAASFROVSCLELYARVQLQC 146

ID AAB10640 standard; Protein; 354 AA.  
 XX  
 XX AAB10640;  
 XX  
 XX 19-JAN-2001 (first entry)  
 XX  
 XX Human VEGF-X protein for expression in Baculovirus/insect cell systems.  
 XX  
 XX VEGF-X; vascular endothelial growth factor; human; vulnery; cytostatic;  
 XX antineumatic; antiarthritic; antiprotic; antidiabetic; treatment;  
 XX angiogenesis regulator; vascularization regulator; cancer; psoriasis;  
 XX rheumatoid arthritis; diabetic retinopathy; blood vessel; organ repair;  
 XX tissue regeneration; tissue repair; wound; dermal ulcer; pressure sore;  
 XX venous sore; diabetic ulcer; burns; skin graft growth.  
 XX  
 XX Homo sapiens.  
 XX  
 XX WO200037641-A2.  
 XX  
 XX 29-JUN-2000.  
 XX  
 XX 21-DEC-1999; 99WO-US30503.  
 XX  
 XX 22-DEC-1998; 98GB-0028377.  
 XX 18-MAR-1999; 99US-0124967.  
 XX 08-NOV-1999; 99US-0164131.  
 XX  
 XX (JANC ) JANSSEN PHARM NV.  
 XX  
 XX Gordon RD, Sprengel JJ, Yon JR, Dijkmans JH, Gosiawska A;  
 XX Dhanaraj SN, Xu J;  
 XX  
 XX WPI: 2000-442669/38.  
 XX N-PSDB; AAA17984.  
 XX  
 XX New vascular endothelial growth factor protein, useful for treating or  
 XX preventing diseases associated with inappropriate angiogenesis activity  
 XX such as cancer, rheumatoid arthritis, psoriasis and wounds -  
 XX  
 XX Disclosure; Fig 20; 127pp; English.  
 XX  
 XX This invention describes a novel vascular endothelial growth factor-X  
 XX (VEGF-X) protein (Ia) and its encoding polynucleotide (IIa) which has  
 XX vulnery, cytostatic, antineumatic, antiarthritic, antipsoriatic and  
 XX antidiabetic activity and acts as an angiogenesis and vascularization  
 XX regulator. An antisense molecule of the invention is useful for treating  
 XX or preventing cancer, rheumatoid arthritis, psoriasis and diabetic  
 XX retinopathy by inhibiting angiogenic activity or inappropriate  
 XX vascularization including formation and proliferation of new blood  
 XX vessels, growth and development of tissues, tissue regeneration and organ  
 XX and tissue repair in a subject. The products of the invention are useful  
 XX for preparing medicaments for treating wounds such as dermal ulcers,  
 XX pressure sores, venous sores, diabetic ulcers and burns and to promote  
 XX skin graft growth, tissue repair, proliferation of new blood vessels,  
 XX tissue regeneration and organ repair by promoting angiogenic activity or  
 XX vascularization. This sequence represents a human VEGF-X protein which  
 XX can be expressed in Baculovirus/insect cell systems and which is  
 XX described in the method of the invention.  
 XX  
 XX Sequence 354 AA;  
 XX  
 QY Query Match 11.2%; Score 106.5; DB 21; Length 354;  
 DB Best Local Similarity 24.4%; Pred. No. 0.0028;  
 Matches 44; Conservative 14; Mismatches 41; Indels 81; Gaps 6;  
 QY 1 MKFLVNVAVLVMVYISYIYARLPDGTGAGEDALPMSKSTAKHPFQIEDNRCYIDNGK 60  
 DB 1 MKFLVNVAVLVMVYISYIYA-----DPESHHHHHSNN----- 35  
 QY 61 LPARGSIVGMRSFVDDPKADYGVGE-----NLVYHADDFEPPGSLKMNVR 109  
 DB 36 -----SSKFPSSNNKQGVODPOHERITVTSGSIHSP-----R 71

DB 110 NLPWMEIFELALRLVGLGVNLKRCVEPLRDYSSAYMMNMOMGAWGQCELLAPDEPED 169

YY 72 FPHTRYFRNTVLWRKLVAEENWIQLT-----FDERPGL-----EDPED 110

RESULT 7  
AAB10641  
ID AAB10641 standard; Protein; 354 AA.

AC AAB10641;  
XX XX  
DT 19-JAN-2001 (first entry)  
XX XX  
DE Human VEGF-X protein for expression in E. coli systems.  
KW VEGF-X; vascular endothelial growth factor; human; vulnery; cytosstatic;  
KM antirheumatic; antiarthritic; antipsoriatic; antidiabetic; treatment;  
KW angiogenesis regulator; vascularization regulator; cancer; psoriasis;  
KM rheumatoid arthritis; diabetic retinopathy; blood vessel; organ repair;  
KM tissue regeneration; tissue repair; wound; dermal ulcer; pressure sore;  
XX venous sore; diabetic ulcer; burns; skin graft growth.  
XX XX  
OS Homo sapiens.  
PN MO200037641-A2.  
PD 29-JUN-2000.  
XX XX  
PF 21-DEC-1999; 99WO-US30503.  
XX XX  
PR 22-DEC-1998; 98GB-0028377.  
PR 18-MAR-1999; 99US-0124967.  
PR 08-NOV-1999; 99US-0164131.  
XX XX  
PA (JANC ) JANSEN PHARM NV.  
PI Gordon RD, Sprengel JJ, Yon JR, Dijkmans JH, Gostiewska A;  
PI Dhanaaraj SN, Xu J;  
XX XX  
DR WPI; 2000-442669/38.  
XX N-PSDB; AAA71985.  
PT New vascular endothelial growth factor protein, useful for treating or  
PT preventing diseases associated with inappropriate angiogenesis actively  
PT such as cancer, rheumatoid arthritis, psoriasis and wounds -  
XX XX  
PS Disclosure; Fig 21, 127pp; English.  
XX XX  
CC This invention describes a novel vascular endothelial growth factor-X  
CC (VEGF-X) protein (Ia) and its encoding polynucleotide (IIa) which has  
CC vulnary, cytostatic, antirheumatic, antiarthritic, antipsoriatic and  
CC antidiabetic activity and acts as an angiogenesis and vascularization  
CC regulator. An antisense molecule of the invention is useful for treating  
CC or preventing cancer, rheumatoid arthritis, psoriasis and diabetic  
CC retinopathy by inhibiting angiogenic activity or inappropriate  
CC vascularization including formation and proliferation of new blood  
CC vessels, growth and development of tissues, tissue regeneration and organ  
CC and tissue repair in a subject. The products of the invention are useful  
CC for preparing medicaments for treating wounds such as dermal ulcers,  
CC pressure sores, venous sores, diabetic ulcers and burns and to promote  
CC skin graft growth, tissue repair, proliferation of new blood vessels,  
CC tissue regeneration and organ repair by promoting angiogenic activity or  
CC vascularization. This sequence represents a human VEGF-X protein which  
CC can be expressed in E. coli systems and which is described in the method  
CC of the invention.  
XX XX  
SQ Sequence 354 AA;

Query Match 11.2%; Score 106.5; DB 21; Length 354;  
Best Local Similarity 24.4%; Pred. No. 0.0028;  
Matches 44; Conservative 14; Mismatches 41; Indels 81; Gaps 6

1 MKPLVANVALVFNVNVIYSIYARLPGDITAGEDALPMKSTAKMHPFOIEDNRCTIDNGK 60

[illegible]



AAB72437  
 ID AAB72437 standard; Peptide; 21 AA.  
 XX  
 AC AAB72437;  
 XX  
 DT 08-MAY-2001 (first entry)  
 XX  
 DE Melittin signal peptide.  
 XX  
 KM UGTF; UDP-glucose:glycoprotein glucosyltransferase; enzyme; honeybee;  
 KM melittin signal peptide.  
 XX  
 OS Apis mellifica.  
 XX  
 PN WO200112845-A1.  
 XX  
 PD 22-FEB-2001.  
 XX  
 PF 27-JUL-2000; 2000WO-CA00883.  
 XX  
 PR 18-AUG-1999; 99US-0376330.  
 XX  
 PA (CANA ) NAT RES COUNCIL CANADA.  
 PI  
 PS Tessier DC, Dignard D, Bergeron JMM, Thomas DY;  
 PI  
 DR WPI; 2001-218358/22.  
 DR N-PSDB; AAF60733.  
 PT  
 PT Determining the effect of a test sample on UDP-glucose:glycoprotein  
 PT glucosyltransferase (UGTF), useful for measuring UGTF activity,  
 PT comprises exposing an acceptor substrate for UGTF to a labeled donor in  
 PT the presence of UGTF -  
 XX  
 PS Example 1; Fig 8; 95pp; English.  
 XX  
 CC The present invention relates to a method for determining the effect of a  
 CC test sample on UDP-glucose:glycoprotein glucosyltransferase (UGTF)  
 CC activity. The method comprises exposing an acceptor substrate for UGTF to  
 CC a labelled donor in the presence of the test sample and UGTF. The method  
 CC is useful for determining UGTF activity. In particular, the method is  
 CC useful in glucosyltransferase assay and kinetics measurement for  
 CC determining UGTF activity. UGTF is a soluble enzyme of the endoplasmic  
 CC reticulum which catalyses the addition of a glucose residue onto  
 CC asparagine-linked oligosaccharides, which are present on incorrectly  
 CC folded glycoproteins. The present sequence is the honeybee melittin  
 CC signal peptide. This sequence was used in the construction of an  
 CC expression vector for rat UGTF (see AAF60732 and AAB72436).  
 CC  
 SQ Sequence 21 AA;  
 Query Match 10.8%; Score 102; DB 22; Length 21;  
 Best Local Similarity 100.0%; Pred. No. 0.00019;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MKELVNVALLVFMVVVYSIYA 21  
 DB 1 MKELVNVALLVFMVVVYSIYA 21  
 RESULT 12  
 ABB31292  
 ID ABB31292 standard; Peptide; 21 AA.  
 XX  
 AC ABB31292;  
 XX  
 DT 21-OCT-2002 (first entry)  
 XX  
 DE Heterologous signal sequence #1.  
 XX  
 KM Signal sequence; ligand identification; hydrophobic protein;  
 KM transmembrane protein; monotypic membrane protein;  
 KM polytopic membrane protein; pump; channel; receptor kinase;

KM G protein-coupled receptor; transporter protein.  
 XX  
 OS Synthetic.  
 XX  
 PN WO200257792-A2.  
 XX  
 PD 25-JUL-2002.  
 XX  
 PF 19-DEC-2001; 2001WO-US50088.  
 XX  
 PR 29-DEC-2000; 2000US-258970P.  
 XX  
 PA (NEOG-) NEOGENESIS PHARM INC.  
 PI  
 PS Felisch JS, Annie DA, Kalghatgi K, Nash HM;  
 PI  
 DR WPI; 2002-599728/64.  
 XX  
 PT Identifying ligand for hydrophobic protein based on affinity selection  
 PT which can operate in the presence of amphiphile without regard to the  
 PT specific biological function of hydrophobic target protein -  
 XX  
 PS Claim 32; Page 80; 97pp; English.  
 XX  
 CC This invention relates to a novel method for identifying a ligand for a  
 CC hydrophobic protein. The method comprises selecting a ligand molecule by  
 CC affinity selection by exposing a hydrophobic target protein bound by an  
 CC amphiphile to a multiplicity of molecules to promote formation of at  
 CC least a complex between the hydrophobic target protein and the ligand  
 CC molecule, separating the complex from the unbound molecules, and  
 CC identifying the ligand molecule. The method of the invention is  
 CC useful for identifying a ligand for hydrophobic protein such as a  
 CC membrane, integral membrane, transmembrane, monotypic or polytopic  
 CC membrane, pump, channel, receptor kinase, G protein-coupled receptor,  
 CC or transporter protein, or membrane-associated enzyme, or Myc tag-EB  
 CC tag-human m2 mACHR, flag tag-human beta2 adrenergic receptor-EB tag,  
 CC human neurokinin 3 receptor-HSV tag-Myc tag, flag tag-human m1 mACHR-EB  
 CC tag, and rat m3 mACHR-HSV tag-Octahis tag. The ligand identified by the  
 CC method of the invention is useful for the development of novel  
 CC medicines and medicinal diagnostics. The present sequence represents  
 CC a heterologous signal sequence peptide used in the method of the  
 CC invention.  
 CC  
 SQ Sequence 21 AA;  
 Query Match 10.8%; Score 102; DB 23; Length 21;  
 Best Local Similarity 100.0%; Pred. No. 0.00019;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MKELVNVALLVFMVVVYSIYA 21  
 DB 1 MKELVNVALLVFMVVVYSIYA 21  
 RESULT 13  
 ABB10034  
 ID ABB10034 standard; Protein; 21 AA.  
 XX  
 AC ABB10034;  
 XX  
 DT 08-OCT-2002 (first entry)  
 XX  
 DE Honeybee melittin signal sequence.  
 XX  
 KM Honeybee; melittin; signal sequence; immunoglobulin; B-cell lymphoma;  
 KM B-cell mediated pathology; cytostatic; immunosuppressive; dermatological;  
 KM antiinflammatory; neuroprotective; antidiabetic; antithyroid;  
 KM autoimmune disease.  
 XX  
 OS Apis mellifica.  
 OS  
 PN WO200213862-A2.  
 XX

PD 21-FEB-2002.  
XX  
PF 10-AUG-2001; 2001MO-US25204.  
XX  
PR 11-AUG-2000; 2000US-224722P.  
XX  
PR 11-AUG-2000; 2000US-224723P.  
XX  
PR 23-MAR-2001; 2001US-279079P.  
XX  
PA (FAVR-) FAVRILE INC.  
XX  
PI Gold DP, Shopes RJ,  
XX  
PI WPI; 2002-280742/32.  
XX  
DR N-PSDB; ABL54969.  
XX  
PT Composition for altering B-cell mediated pathology, has a chimeric  
XX  
PT protein having portion of variable region of heavy chain or light chain  
XX  
PT linked to portion constant region associated with patient B cell clone  
XX  
PS Claim 19; Page 17; 100pp; English.  
XX  
CC The sequence represents the honeybee melittin signal sequence, used in  
XX  
CC the invention to maximise secretion of heterologous proteins from insect  
XX  
CC cells. The invention relates to a novel composition for altering a B-cell  
XX  
CC mediated pathology in a patient. The composition contains a chimeric  
XX  
CC protein comprising at least a portion of a variable region of heavy chain  
XX  
CC or light chain (VH or VL) linked to at least a portion of an  
XX  
CC immunoglobulin constant region, where VH or VL region is associated with  
XX  
CC a B cell clone from the patient having the B cell mediated pathology. The  
XX  
CC composition of the invention has cytostatic, immunosuppressive,  
XX  
CC dermatological, antiinflammatory, neuroprotective, antidiabetic, and  
XX  
CC antitumor activity. The composition is a vaccine useful for altering a  
XX  
CC B cell mediated pathology. This includes B cell lymphoma e.g.  
XX  
CC non-Hodgkins lymphoma, refractory low grade or follicular B cell  
XX  
CC lymphoma; autoimmune disease e.g. multiple sclerosis, systemic lupus  
XX  
CC erythematosus, anti-Hu associated paraneoplastic neurological syndrome,  
XX  
CC autoimmune hepatitis, Type I diabetes, autoimmune thyroiditis and  
XX  
CC scleroderma. The pathology is treated by administering the composition to  
XX  
CC the patient, preferably with a cytokine e.g. granulocyte-macrophage  
XX  
CC colony stimulating factor (GM-CSF) or chemokine e.g. monocyte chemoattractant  
XX  
CC protein 3 (MCP 3).  
XX  
SQ Sequence 21 AA;  
XX  
Query Match 10.8%; Score 102; DB 23; Length 21;  
XX  
Best Local Similarity 100.0%; Pred. No. 0.00019;  
XX  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 MKFLVNVVALVFMVVYISYIYA 21  
DB 1 MKFLVNVVALVFMVVYISYIYA 21  
XX  
RESULT 14  
AAU76918  
ID AAU76918 standard; Peptide: 21 AA.  
XX  
AC AAU76918;  
XX  
DT 05-JUN-2002 (first entry)  
XX  
DE Honeybee melittin signal sequence.  
XX  
XX Honeybee; signal sequence; rheumatoid arthritis; melittin;  
XX  
XX Vbeta; Valpha; T cell receptor; TCR; cytostatic; neuroprotective;  
XX  
XX antidiabetic; dermatological; antiinflammatory; immunosuppressive;  
XX  
XX antiinflammatory; antitumor; antitumor; antitumor;  
XX  
XX autoimmune disease; multiple sclerosis; systemic lupus erythematosus;  
XX  
XX diabetes; inflammatory bowel disease; myasthenia gravis; thyroiditis.  
XX  
OS Apis mellifica.  
XX

PN WO200213861-A2.  
XX  
PD 21-FEB-2002.  
XX  
XX  
PF 10-AUG-2001; 2001MO-US25203.  
XX  
XX  
PR 11-AUG-2000; 2000US-224722P.  
XX  
PR 11-AUG-2000; 2000US-224723P.  
XX  
PR 01-FEB-2001; 2001US-266133P.  
XX  
PA (FAVR-) FAVRILE INC.  
XX  
XX Gold DP, Shopes RJ,  
XX  
PI WPI; 2002-241838/29.  
XX  
DR N-PSDB; ABK10526.  
XX  
PT Altering T cell mediated pathology in patient, involves administering  
XX  
PT composition having chimeric protein which has portion of Vbeta/Valpha  
XX  
PT region of T cell receptor and portion of immunoglobulin constant region  
XX  
PT  
XX  
PS Disclosure; Page 8; 110pp; English.  
XX  
XX This invention relates to a method for altering a T cell mediated  
XX  
XX pathology in a patient, the method involves administering a composition  
XX  
XX comprising a chimeric protein which comprises at least a portion of a  
XX  
XX Vbeta or Valpha region of a T cell receptor (TCR), and at least a  
XX  
XX portion of an immunoglobulin (Ig) constant region, where the Vbeta or  
XX  
XX Valpha region is associated with a particular TCR from a T cell from the  
XX  
XX patient having T cell mediated pathology. The method may have  
XX  
XX cytostatic, neuroprotective, antidiabetic, dermatological,  
XX  
XX antitumor activity. The composition of the invention is useful for  
XX  
XX altering a T cell mediated pathology in a patient, where the T cell  
XX  
XX mediated pathology is T cell lymphoma, or an autoimmune disease selected  
XX  
XX from multiple sclerosis, systemic lupus erythematosus, diabetes,  
XX  
XX inflammatory bowel disease, myasthenia gravis, rheumatoid arthritis, and  
XX  
XX thyroiditis. The present sequence represents the honeybee melittin  
XX  
XX secretory signal sequence, this sequence was used as a leader sequence  
XX  
XX in to maximise the secretion of heterologous proteins from transfected  
XX  
XX insect cells.  
XX  
SQ Sequence 21 AA;  
XX  
Query Match 10.8%; Score 102; DB 23; Length 21;  
XX  
Best Local Similarity 100.0%; Pred. No. 0.00019;  
XX  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 MKFLVNVVALVFMVVYISYIYA 21  
DB 1 MKFLVNVVALVFMVVYISYIYA 21  
XX  
RESULT 15  
AAM52328  
ID AAM52328 standard; Peptide: 21 AA.  
XX  
AC AAM52328;  
XX  
DT 28-JAN-2002 (first entry)  
XX  
DE Honeybee melittin signal peptide #1.  
XX  
XX Honeybee; signal peptide; melittin.  
XX  
XX Apis mellifica.  
XX  
OS JP2001258565-A.  
XX  
XX 25-SEP-2001.  
XX  
PD 21-MAR-2000; 2000JP-0078897.  
XX  
PF







GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: January 22, 2004, 11:37:25 ; Search time 26.5802 Seconds

(without alignments)  
313.155 Million cell updates/sec

Title: US-09-890-806-5

Perfect score: 947  
Sequence: 1 MKFLVVALVFMVVVYSIY.....GQELAPEDPEDSALLIEDPV 177

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_41.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	740	78.1	199	1 US02_HCMVA	P09713 human cytom
2	125	13.2	186	1 US02_HCMVA	P09712 human cytom
3	102	10.8	70	1 MEL_APIME	P01501 apis mellif
4	102	10.8	70	1 MEL_VESMC	P59262 vespula mac
5	99	10.5	393	1 VGLD_HSV1H	P06476 herpes simp
6	99	10.5	394	1 VGLD_HSV1H	P06991 herpes simp
7	99	10.5	394	1 VGLD_HSV1A	P36318 herpes simp
8	99	10.5	394	1 VGLD_HSV1P	P59260 apis cerana
9	98.5	10.4	70	1 MEL_APIPC	P59260 apis cerana
10	98.5	10.4	70	1 MEL_POLHE	P59261 polistes be
11	96.5	10.2	1409	1 COP1_PROME	P04146 drosophila
12	91	9.6	393	1 VGLD_HSV2	P03172 herpes simp
13	81	8.6	410	1 REPI_ZYGRO	P13778 zygosacchar
14	80	8.4	703	1 CDGT_BACS2	P31746 bacillus sp
15	78.5	8.3	368	1 YJHT_ECOLI	P39371 escherichia
16	77.5	8.2	248	1 GRAT_GYXN	O12705 xestia c-ni
17	77	8.1	272	1 GPH1_PSEAE	O98586 pseudomonas
18	76.5	8.1	248	1 GRAT_GYXN	P06503 trichoplusi
19	76.5	8.1	316	1 AKBA_HUMAN	P06218 homo sapien
20	75	7.9	315	1 ALDR_PIG	P80216 sus scrofa
21	75	7.9	620	1 BCHD_CHLTE	O935W0 chlorobium
22	74.5	7.9	248	1 GRAT_GYXN	O91B74 adoxophyes
23	74	7.8	261	1 ZNBE_ECOLI	P39832 escherichia
24	73	7.7	463	1 UHPF_SALTY	P27670 salmonella
25	73	7.7	1620	1 HEDI_HUMAN	O9UL88 homo sapien
26	72.5	7.7	446	1 LAMB_ECOL6	O8CV14 escherichia
27	72.5	7.7	466	1 HEMN_SYNY3	P74132 synchocyst
28	72.5	7.7	619	1 BCHD_CHLVI	O50133 chlorobium
29	72	7.6	192	1 VIP_HVIB5	P04588 human immun
30	72	7.6	704	1 CDGT_BACCH	P27036 bacillus oh
31	71.5	7.6	1036	1 ATHT_RAT	P54708 rattus norv
32	71.5	7.6	1266	1 NGCA_CHICK	O03696 gallus gall
33	71	7.5	308	1 AMID_STRPN	P18794 streptococc

34	71	7.5	491	1 PBP_BACSU	P39844 bacillus su
35	70.5	7.4	361	1 GLNA_PANAR	O04831 panulirus a
36	70.5	7.4	607	1 LEPA_STRPN	O976K5 streptococc
37	70.5	7.4	811	1 MUTS_THBAQ	O56215 thermus acu
38	70	7.4	446	1 LAMB_ECOLI	P02943 escherichia
39	70	7.4	859	1 OBP_HSVBC	P52377 bovine herp
40	69.5	7.3	349	1 US10_HCMVA	P09706 human cytom
41	69.5	7.3	446	1 LAMB_ECOS7	O8X5W7 escherichia
42	69.5	7.3	562	1 GUNI_ACICE	P54583 acidothermu
43	69.5	7.3	878	1 SVV_METUA	O58413 methanococc
44	69.5	7.3	1059	1 CERU_RAT	P13635 rattus norv
45	69	7.3	286	1 ATRPS_SCHPO	P87127 schizosacch

## ALIGNMENTS

RESULT 1	US02_HCMVA	STANDARD;	PRT;	199 AA.
ID	US02_HCMVA			
AC	P09713;			
DT	01-MAR-1989 (Rel. 10, Created)			
DT	01-MAR-1989 (Rel. 10, Last sequence update)			
DT	15-SEP-2003 (Rel. 42, Last annotation update)			
DE	Hypothetical protein HQF2.			
GN	US2.			
OS	Human cytomegalovirus (strain AD169).			
OC	viruses; dsDNA viruses, no RNA stage; Herpesviridae;			
OC	Betaherpesvirinae; Cytomegalovirus.			
OX	NCBI_TaxID=10360;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=87169717; PubMed=3031311;			
RA	Weston K., Barrell B.G.;			
RT	"Sequence of the short unique region, short repeats, and part of the			
RT	long repeats of human cytomegalovirus.";			
RL	J. Mol. Biol. 192:177-208(1986).			
RN	[2]			
RP	COMPLETE GENOME.			
RX	MEDLINE=90269039; PubMed=2161319;			
RA	Chee M.S., Bankier A.T., Beck S., Bohni R., Brown C.M., Cerny R.,			
RA	Horsnell T., Hutchinson C.A. III, Kouzarides T., Martignetti J.A.,			
RA	Predlie E., Satchwell S.C., Tomlinson P., Weston K.M., Barrell B.G.;			
RT	"Analysis of the protein-coding content of the sequence of human			
RT	cytomegalovirus strain AD169."			
RL	Curr. Top. Microbiol. Immunol. 154:125-169(1990).			
CC	-1- SIMILARITY: BELONGS TO THE US2 FAMILY.			
CC				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>			
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).			
CC				
DR	EMBL; X17403; CA45113.1; -			
DR	EMBL; X04650; CAB37096.1; -			
DR	PIR; E26078; Q0BEC5.			
KW	PDB; 1IM3; 27-JUN-01.			
HM	Hypothetical protein; 3D-structure.			
FT	CARBOHYD 68 68 N-LINKED (GLCNAC. . .) (POTENTIAL).			
FT	CARBOHYD 172 172 N-LINKED (GLCNAC. . .) (POTENTIAL).			
FT	CARBOHYD 188 188 N-LINKED (GLCNAC. . .) (POTENTIAL).			
SQ	SEQUENCE 199 AA; 23111 MW; 4DD2DF3D692393F3 CRC64;			
Query Match	78.1%; Score 740; DB 1; Length 199;			
Best local Similarity	100.0%; Pred. No. 3e-65;			
Matches 136; Conservative	0; Mismatches 0; Indels 0; Gaps 0;			
QY	22 RLPDGTAKGEDALRPWKTATAPWFOIEDNRCYIDNGKLFARGSTVGNMSPFVDPKAD 81			
DB	20 RLPDGTAKGEDALRPWKTATAPWFOIEDNRCYIDNGKLFARGSTVGNMSPFVDPKAD 79			

QY 82 YGSGENLVYHADVEFPGESLKMNVRLNDVMPFETTLALRLVLOQDVYWLRCVPELRV 141  
 DB 80 YGSGENLVYHADVEFPGESLKMNVRLNDVMPFETTLALRLVLOQDVYWLRCVPELRV 139  
 QY 142 DYTSSAYMMNMXYGMV 157  
 DB 140 DYTSSAYMMNMXYGMV 155

RESULT 2  
 US03 HCMVA STANDARD; PRT; 186 AA.  
 ID US03 HCMVA  
 AC P09712;  
 DT 01-MAR-1989 (Rel. 10, Created)  
 DT 01-MAR-1989 (Rel. 10, Last sequence update)  
 DT 01-FEB-1991 (Rel. 17, Last annotation update)  
 DE HQLF1 protein (E glycoprotein).  
 GN US3.  
 OS Human cytomegalovirus (strain AD169).  
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
 OC Betaherpesvirinae; Cytomegalovirus.  
 OC NCBI\_TaxID=10360;  
 RX MEDLINE=87169717; PubMed=3031311;  
 RA Weston K., Barrell B.G.;  
 RT "Sequence of the short unique region, short repeats, and part of the  
 RT long repeats of human cytomegalovirus.";  
 RL J. Mol. Biol. 192:177-208(1986).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=87169717; PubMed=3031311;  
 RA Weston K., Barrell B.G.;  
 RT "An enhancer element in the short unique region of human  
 RT cytomegalovirus regulates the production of a group of abundant  
 RT immediate early transcripts.";  
 RL Virology 162:406-416(1988).  
 RN [3]  
 RP COMPLETE GENOME.  
 RX MEDLINE=90269039; PubMed=2161319;  
 RA Chee M.S., Bankier A.T., Beck S., Bohni R., Brown C.M., Cerny R.,  
 RA Hornslell T., Hutchison C.A. III, Kouzarides T., Martignetti J.A.,  
 RA Preddie B., Satchwell S.C., Tomlinson P., Weston K.M., Barrell B.G.;  
 RT "Analysis of the protein-coding content of the sequence of human  
 RT cytomegalovirus strain AD169.";  
 RL Curr. Top. Microbiol. Immunol. 154:125-169(1990).  
 CC CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: X17403; CA35314.1; -  
 DR EMBL: X04650; CA37097.1; -  
 DR EMBL: M18921; AAA45958.1; -  
 DR PIR: P26078; Q0BEC6.  
 KW Early protein; Glycoprotein.  
 FT CARBOHYD 60  
 FT SEQUENCE 186 AA; 21574 MW; 671753CLAA75920D CRC64; .

Query Match 13.2%; Score 125; DB 1; Length 186;  
 Best Local Similarity 23.9%; Pred. No. 3.8e-05;  
 Matches 39; Conservative 36; Mismatches 56; Indels 32; Gaps 8;

QY 10 VPMVVVSYIYARPDGTTXAGDALPWPWKSTAGHPQIDNRCYIDNGTLFARGSIYV 69  
 DB 4 VLVIAIYAVLFRLADSVPRPLDVVV-----SEISAHFRVENOCWFPMGLYKGRSG 59  
 QY 70 NMSRFVDPKADYGVGENTL-----YHADVEFPGESLKMNV-----RLNDVM 114

DB 60 NFEKHF---VAVGVISQSYMDRLQVSGEQYHND-----RGAYFFPMNIGGHPVHTVDMV 112  
 QY 115 PIFETTLALRLVLOQD-VIWLRCVPELRVDTSSAYMMNMXYGMV 156  
 DB 113 DI--ILSTR---WGDPKYAACVPOVRMDYSSQITINWYLRSM 150

RESULT 3  
 MEL\_APIME STANDARD; PRT; 70 AA.  
 ID MEL\_APIME  
 AC P01501; P01503;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 21-JUL-1986 (Rel. 01, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Melittin precursor (Allergen Api m 3) (Api m III).  
 GN MELT.  
 OS Apis mellifera (Honeybee), and  
 OS Apis cerana (Indian honeybee).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Hymenoptera; Apocrita; Aculeata; Apoidea;  
 OC Apidae; Apis.  
 OC NCBI\_TaxID=7460, 7461;  
 RX MEDLINE=83287387; PubMed=6309516;  
 RA Vlasak R., Unger-Ullmann C., Kreil G., Fritsch A.-M.;  
 RT "Nucleotide sequence of cloned cDNA coding for honeybee  
 RT prepro-melittin.";  
 RL Eur. J. Biochem. 135:123-126(1983).  
 RN [2]  
 RP SEQUENCE OF 44-69 (MELITTIN 1 AND 2).  
 RX MEDLINE=68327913; PubMed=5592400;  
 RA Habermann E., Jentsch U.;  
 RT "Sequence analysis of melittin from tryptic and peptic degradation  
 RT products.";  
 RL Hoppe-Seyler's Z. Physiol. Chem. 348:37-50(1967).  
 RN [3]  
 RP SYNTHESIS.  
 RX MEDLINE=72098668; PubMed=5139482;  
 RA Schroeder E., Luebke K., Lehmann M., Beetz I.;  
 RT "Haemolytic activity and action on the surface tension of aqueous  
 RT solutions of synthetic melittins and their derivatives.";  
 RL Experientia 27:764-765(1971).  
 RN [4]  
 RP SYNTHESIS.  
 RX MEDLINE=72098669; PubMed=5139483;  
 RA Luebke K., Matthes S., Kloss G.;  
 RT "Isolation and structure of N 1-formyl melittin.";  
 RL Experientia 27:765-767(1971).  
 RN [5]  
 RP SEQUENCE OF 44-69.  
 RX SPECIES=A.cerana;  
 RC Kreil G.;  
 RT "Structure of melittin isolated from two species of honey bees.";  
 RL FEBS Lett. 33:241-244(1973).  
 RN [6]  
 RP X-RAY CRYSTALLOGRAPHY.  
 RX SPECIES=A.mellifera;  
 RC MEDLINE=82189959; PubMed=7076662;  
 RA Terwilliger T.C., Eisenberg D.;  
 RT "The structure of melittin. II. Interpretation of the structure.";  
 RL J. Biol. Chem. 257:6016-6022(1982).  
 RN [7]  
 RP STRUCTURE BY NMR OF 44-69.  
 RX SPECIES=A.mellifera;  
 RC Barnham K.J., Hewish D., Werkmeister J., Curtain C., Kirkpatrick A.,  
 RA Bartone N., Norton R., Rivett D.;  
 RT Submitted (JUN-1998) to the PDB data bank.

RN [8]  
 RP REVIEW.  
 RX MEDLINE=90254148; PubMed=2187536;  
 RA Dempsey C.E.;  
 RT "The actions of melittin on membranes";  
 RL Biochim. Biophys. Acta 1031:143-161(1990).  
 CC -1- FUNCTION: Main toxin of bee venom with strong hemolytic activity.  
 CC integrates into cell membranes and has multiple effects, probably,  
 CC as a result of its interaction with negatively charged  
 CC phospholipids. It inhibits well known transport pumps such as the  
 CC Na(+)-K(+)-ATPase and the H(+)-K(+)-ATPase. Increases the  
 CC permeability of cell membranes to ions, particularly Na+ and  
 CC indirectly Ca(2+), because of the Na(+)-Ca(2+)-exchange.  
 CC -1- SUBUNIT: Monomer and homotrimer.  
 CC -1- SUBCELLULAR LOCATION: Secreted.  
 CC -1- TISSUE SPECIFICITY: Expressed by the venom gland.  
 CC -1- MISCELLANEOUS: N-formyl-melittin major has 80% of the activity of  
 CC melittin.  
 CC -1- SIMILARITY: BELONGS TO THE MELITTIN FAMILY.  
 CC -1- DATABASE: NAME=Protein Spotlight;  
 CC NOTE=Issue 12 of July 2001;  
 CC WWW=http://www.expasy.org/spotlight/articles/split012.html".  
 -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 -----  
 DR EMBL: X02007; CAA26038.1; -  
 DR PIR: A01762; MEHBL.  
 DR PIR: A91133; MEHBL.  
 DR PDB: 2MLT; 15-JUL-92.  
 DR PDB: 1BH1; 06-JAN-99.  
 DR InterPro: IPR002116; Melittin.  
 DR Pfam: PF01372; Melittin; 1.  
 DR ProDom: PD014636; Melittin; 1.  
 DR Cytolsys: Hemolysins; Toxin; Allergen; Signal; Amidation; Formylation;  
 KW 3D-structure.  
 FT SIGNAL 1 21  
 FT PROPEP 22 43  
 FT CHAIN 44 69  
 FT MOD\_RES 44 44  
 FT MOD\_RES 69 69  
 FT VARIANT 64 64  
 FT VARIANT 67 70  
 FT HELIX 45 53  
 FT TURN 54 54  
 FT HELIX 55 68  
 SQ SEQUENCE 70 AA; 7585 MW; 607F52C091C23BB6 CRC64;  
 -----  
 QY 1 MKFLVNVALVFWMVYISYIYA 21  
 DB 1 MKFLVNVALVFWMVYISYIYA 21  
 -----  
 RESULT 4  
 ID MEL\_VESMC STANDARD; PRT; 70 AA.  
 AC P59762;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Melittin precursor.  
 DE MELT.

OS Vespaula maculifrons (Eastern yellow jacket) (Wasp).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Hymenoptera; Apocrita; Aculeata; Vespoidea;  
 OC Vespidae; Vespinae; Vespaula.  
 OX NCBI\_TaxID=7453;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Venom gland;  
 RA Shi W.J., Zhang S.F., Zhang C.X., Cheng J.A.;  
 RT "Cloning and sequencing of cDNA coding for pre-melittin of Polistes  
 RT hebraeus, Vespa magnifica and Vespaula maculifrons";  
 RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: Has strong hemolytic activity. Integrates into cell  
 CC membranes and has multiple effects, probably, as a result of its  
 CC interaction with negatively charged phospholipids. It inhibits  
 CC well known transport pumps such as the Na(+)-K(+)-ATPase and the  
 CC H(+)-K(+)-ATPase. Increases the permeability of cell membranes to  
 CC ions, particularly Na+ and indirectly Ca(2+), because of the  
 CC Na(+)-Ca(2+)-exchange (By similarity).  
 CC -1- SUBUNIT: Monomer and homotrimer (By similarity).  
 CC -1- SUBCELLULAR LOCATION: Secreted (By similarity).  
 CC -1- TISSUE SPECIFICITY: Expressed by the venom gland.  
 CC -1- SIMILARITY: BELONGS TO THE MELITTIN FAMILY.  
 -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 -----  
 DR EMBL: AF487911; AA012205.1; -  
 DR Pfam: PF01372; Melittin; 1.  
 DR ProDom: PD014636; Melittin; 1.  
 DR Cytolsys: Hemolysins; Toxin; Signal; Amidation.  
 KW SIGNAL 1 21  
 FT PROPEP 22 43  
 FT CHAIN 44 69  
 FT MOD\_RES 69 69  
 FT MOD\_RES 69 69  
 FT VARIANT 64 64  
 FT VARIANT 67 70  
 FT HELIX 45 53  
 FT TURN 54 54  
 FT HELIX 55 68  
 SQ SEQUENCE 70 AA; 7585 MW; 607F52C091C23BB6 CRC64;  
 -----  
 QY 1 MKFLVNVALVFWMVYISYIYA 21  
 DB 1 MKFLVNVALVFWMVYISYIYA 21  
 -----  
 RESULT 5  
 ID VGD\_HSV1H STANDARD; PRT; 393 AA.  
 AC P06476;  
 DT 01-JAN-1988 (Rel. 06, Created)  
 DT 01-JAN-1988 (Rel. 06, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Glycoprotein D precursor.  
 GN GD OR US6.  
 OS Herpes simplex virus (type 1 / strain HZT).  
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
 OC Alphaherpesvirinae; Simplexvirus.  
 OX NCBI\_TaxID=10305;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=84131549; PubMed=6321120;  
 RA Laasy L.A., Dowdenko D.O.;  
 RT "DNA sequence analysis of the type-common glycoprotein-D genes of  
 RT herpes simplex virus types 1 and 2.";  
 RL DNA 3:23-29(1984).

CC -1- MISCELLANEOUS: THERE ARE SEVEN EXTERNAL GLYCOPROTEINS IN HSV1: GH,  
CC GB, GC, GG, GD, GI, AND GE.  
CC -1- SIMILARITY: BELONGS TO THE HERPESVIRUSES GLYCOPROTEIN D FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: K02372; AAA45786.1; -  
CC DR InterPro: IPR002896; Herpes\_glycop\_D.  
CC DR Pfam: PF01537; Herpes\_glycop\_D; 1.  
CC KW Glycoprotein; Transmembrane; Signal.  
CC FT SIGNAL 1 25 POTENTIAL.  
CC FT CHAIN 26 393 GLYCOPROTEIN D.  
CC FT DOMAIN 26 338 EXTRACELLULAR (POTENTIAL).  
CC FT TRANSMEM 339 363 POTENTIAL.  
CC FT DOMAIN 364 393 CYTOPLASMIC (POTENTIAL).  
CC FT DOMAIN 364 380 ARG/LYS-RICH (HIGHLY BASIC; PROBABLY  
CC SERVES TO ANCHOR THE GLYCOPROTEIN IN  
CC THE MEMBRANE).  
CC FT CARBOHYD 118 118 N-LINKED (GLCNAC. . .) (POTENTIAL).  
CC FT CARBOHYD 145 145 N-LINKED (GLCNAC. . .) (POTENTIAL).  
CC FT CARBOHYD 286 286 N-LINKED (GLCNAC. . .) (POTENTIAL).  
CC SQ SEQUENCE 393 AA; 43368 MW; 3488035BA4B6FE9 CRC64;  
CC -----  
CC Query Match 10.5%; Score 99; DB 1; Length 393;  
CC Best Local Similarity 100.0%; Pred. No. 0.032;  
CC Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
CC -----  
CC QY 159 QPELAPEDPDSALTEDPV 177  
CC |||||  
CC 289 QPELAPEDPDSALTEDPV 307  
CC -----  
CC RESULT 6  
CC ID VGLD\_HSV11 STANDARD; PRT; 394 AA.  
CC AC Q69091; O12544; O12833; P03171;  
CC DT 21-JUL-1986 (Rel. 01, Created)  
CC DT 16-OCT-2001 (Rel. 40, Last sequence update)  
CC DT 16-OCT-2001 (Rel. 40, Last annotation update)  
CC DE Glycoprotein D precursor.  
CC GN GD OR US6.  
CC OS Herpes simplex virus (type 1 / strain 17).  
CC OS Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
CC OC Alphaherpesvirinae; Simplexvirus.  
CC OX NCBI\_TaxID=10299;  
CC RN [1]  
CC RP SEQUENCE FROM N.A.  
CC RX MEDLINE=85160822; PubMed=2984429;  
CC RA McGeoch D.J., Dolan A., Donald S., Rixon F.J.;  
CC RT "Sequence determination and genetic content of the short unique region  
CC in the genome of herpes simplex virus type 1.",  
CC J. Mol. Biol. 181:1-13(1985).  
CC RL [2]  
CC RN REVISIONS.  
CC RA McGeoch D.J.;  
CC RL Submitted (JUN-1989) to the EMBL/Genbank/DBJ databases.  
CC -1- MISCELLANEOUS: THERE ARE SEVEN EXTERNAL GLYCOPROTEINS IN HSV1: GH,  
CC GB, GC, GG, GD, GI, AND GE.  
CC -1- SIMILARITY: BELONGS TO THE HERPESVIRUSES GLYCOPROTEIN D FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

CC EMBL: L00036; AAA96682.1; -  
CC DR EMBL: X14112; CAA32283.1; -  
CC DR PIR: A03730; VGBB17.  
CC DR InterPro: IPR002896; Herpes\_glycop\_D.  
CC DR Pfam: PF01537; Herpes\_glycop\_D; 1.  
CC KW Glycoprotein; Transmembrane; Signal.  
CC FT SIGNAL 1 25 POTENTIAL.  
CC FT CHAIN 26 394 GLYCOPROTEIN D.  
CC FT DOMAIN 26 339 EXTRACELLULAR (POTENTIAL).  
CC FT TRANSMEM 340 364 POTENTIAL.  
CC FT DOMAIN 365 394 ARG/LYS-RICH (HIGHLY BASIC; PROBABLY  
CC SERVES TO ANCHOR THE GLYCOPROTEIN IN  
CC THE MEMBRANE).  
CC FT CARBOHYD 119 119 N-LINKED (GLCNAC. . .) (POTENTIAL).  
CC FT CARBOHYD 146 146 N-LINKED (GLCNAC. . .) (POTENTIAL).  
CC FT CARBOHYD 287 287 N-LINKED (GLCNAC. . .) (POTENTIAL).  
CC SQ SEQUENCE 394 AA; 43346 MW; 94B5217F355F6FC CRC64;  
CC -----  
CC Query Match 10.5%; Score 99; DB 1; Length 394;  
CC Best Local Similarity 100.0%; Pred. No. 0.032;  
CC Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
CC -----  
CC QY 159 QPELAPEDPDSALTEDPV 177  
CC |||||  
CC 290 QPELAPEDPDSALTEDPV 308  
CC -----  
CC RESULT 7  
CC ID VGLD\_HSV1A STANDARD; PRT; 394 AA.  
CC AC P36318;  
CC DT 01-JUN-1994 (Rel. 29, Created)  
CC DT 01-JUN-1994 (Rel. 29, Last sequence update)  
CC DT 16-OCT-2001 (Rel. 40, Last annotation update)  
CC DE Glycoprotein D precursor.  
CC GN GD OR US6.  
CC OS Herpes simplex virus (type 1 / strain Angelotti).  
CC OS Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
CC OC Alphaherpesvirinae; Simplexvirus.  
CC OX NCBI\_TaxID=10301;  
CC RN [1]  
CC RP SEQUENCE FROM N.A.  
CC RX MEDLINE=90324869; PubMed=2165127;  
CC RA Trunfi K.M., Stevens J.G.;  
CC RT "Molecular and biological characterization of a herpes simplex virus  
CC type 1 (HSV-1) neuroinvasiveness gene.",  
CC J. Exp. Med. 172:487-496(1990).  
CC -1- MISCELLANEOUS: THERE ARE SEVEN EXTERNAL GLYCOPROTEINS IN HSV1: GH,  
CC GB, GC, GG, GD, GI, AND GE.  
CC -1- SIMILARITY: BELONGS TO THE HERPESVIRUSES GLYCOPROTEIN D FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: X54361; CAA38245.1; -  
CC DR PIR: A47627; A47627.  
CC DR InterPro: IPR002896; Herpes\_glycop\_D.  
CC DR Pfam: PF01537; Herpes\_glycop\_D; 1.  
CC KW Glycoprotein; Transmembrane; Signal.  
CC FT SIGNAL 1 25 POTENTIAL.  
CC FT CHAIN 26 394 GLYCOPROTEIN D.  
CC FT DOMAIN 26 339 EXTRACELLULAR (POTENTIAL).  
CC FT TRANSMEM 340 364 POTENTIAL.  
CC FT DOMAIN 365 394 ARG/LYS-RICH (HIGHLY BASIC; PROBABLY  
CC SERVES TO ANCHOR THE GLYCOPROTEIN IN  
CC THE MEMBRANE).  
CC FT CARBOHYD 119 119 N-LINKED (GLCNAC. . .) (POTENTIAL).  
CC FT CARBOHYD 146 146 N-LINKED (GLCNAC. . .) (POTENTIAL).  
CC FT CARBOHYD 287 287 N-LINKED (GLCNAC. . .) (POTENTIAL).  
CC SQ SEQUENCE 394 AA; 43346 MW; 94B5217F355F6FC CRC64;  
CC -----  
CC Query Match 10.5%; Score 99; DB 1; Length 394;  
CC Best Local Similarity 100.0%; Pred. No. 0.032;  
CC Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
CC -----  
CC QY 159 QPELAPEDPDSALTEDPV 177  
CC |||||  
CC 290 QPELAPEDPDSALTEDPV 308  
CC -----  
CC RESULT 7  
CC ID VGLD\_HSV1A STANDARD; PRT; 394 AA.  
CC AC P36318;  
CC DT 01-JUN-1994 (Rel. 29, Created)  
CC DT 01-JUN-1994 (Rel. 29, Last sequence update)  
CC DT 16-OCT-2001 (Rel. 40, Last annotation update)  
CC DE Glycoprotein D precursor.  
CC GN GD OR US6.  
CC OS Herpes simplex virus (type 1 / strain Angelotti).  
CC OS Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
CC OC Alphaherpesvirinae; Simplexvirus.  
CC OX NCBI\_TaxID=10301;  
CC RN [1]  
CC RP SEQUENCE FROM N.A.  
CC RX MEDLINE=90324869; PubMed=2165127;  
CC RA Trunfi K.M., Stevens J.G.;  
CC RT "Molecular and biological characterization of a herpes simplex virus  
CC type 1 (HSV-1) neuroinvasiveness gene.",  
CC J. Exp. Med. 172:487-496(1990).  
CC -1- MISCELLANEOUS: THERE ARE SEVEN EXTERNAL GLYCOPROTEINS IN HSV1: GH,  
CC GB, GC, GG, GD, GI, AND GE.  
CC -1- SIMILARITY: BELONGS TO THE HERPESVIRUSES GLYCOPROTEIN D FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

FT THE MEMBRANE).  
 FT CARBOHYD 119 119 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 146 146 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 287 287 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 394 AA; 43303 MW; 47DE3BC79BB28950 CRC64;

Query Match 10.5%; Score 99; DB 1; Length 394;  
 Best Local Similarity 100.0%; Pred. No. 0.032;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 159 QPELAPEDPDSALLDDPV 177  
 DB 290 QPELAPEDPDSALLDDPV 308

## RESULT 8

VGDL HSV1P STANDARD; PRT; 394 AA.  
 AC P57083; P03171;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 21-JUL-1986 (Rel. 01, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Glycoprotein D precursor.  
 GN GD OR US6.  
 OS Herpes simplex virus (type 1 / strain Patton).  
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
 CC Alphaherpesvirinae; Simplexvirus.  
 NCBI\_TaxID=10308;

RP SEQUENCE FROM N.A.  
 RA MEDLINE=83016630; PubMed=6289440;  
 Watson R.J., Weis J.H., Salestrom J.S., Engquist L.W.;  
 RT "Herpes simplex virus type-1 glycoprotein D gene: nucleotide sequence and expression in Escherichia coli.";  
 Science 218:381-384 (1982).  
 CC -1- MISCELLANEOUS: THERE ARE SEVEN EXTERNAL GLYCOPROTEINS IN HSV1: GH,  
 GB, GC, GG, GI, AND GE.  
 CC -1- SIMILARITY: BELONGS TO THE HERPESVIRUSES GLYCOPROTEIN D FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; X02138; AAA26060.1; -;  
 DR EMBL; J02217; AAA45785.1; -;  
 DR PIR; A94268; VGBED1.  
 DR PDB; 1JMA; 26-SEP-01.  
 DR InterPro; IPR002896; Herpes\_glycop.D.  
 DR Pfam; PF01537; Herpes\_glycop.D; 1.  
 KW Glycoprotein; Transmembrane; Signal; 3D-structure.  
 FT SIGNAL 1 25.  
 FT CHAIN 26 394 GLYCOPROTEIN D.  
 FT DOMAIN 26 339 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 340 364 POTENTIAL.  
 FT DOMAIN 365 394 CYTOPLASMIC (POTENTIAL).  
 FT DOMAIN 381 381 ARG/LYS-RICH (HIGHLY BASIC; PROBABLY  
 SERVES TO ANCHOR THE GLYCOPROTEIN IN  
 THE MEMBRANE).  
 FT CARBOHYD 119 119 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 146 146 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 287 287 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 394 AA; 43346 MW; 052AB5F5303D5E CRC64;

Query Match 10.5%; Score 99; DB 1; Length 394;  
 Best Local Similarity 100.0%; Pred. No. 0.032;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 159 QPELAPEDPDSALLDDPV 177

DB 290 QPELAPEDPDSALLDDPV 308

## RESULT 9

MEL\_APICC STANDARD; PRT; 70 AA.  
 ID MEL\_APICC  
 AC P59260;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Melittin precursor.  
 GN MELT.  
 OS Apis cerana cerana (Oriental honeybee),  
 OS Vespa magnifica (Hornet), and  
 OS Vespa velutina nigrithorax (Hornet).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Hymenoptera; Apocrita; Aculeata; Apoidea;  
 OC Apidae; Apis.  
 NCBI\_TaxID=94128, 202807, 202809;

RP SEQUENCE FROM N.A.  
 RA SPECIES-A.c.cerana, V.magnifica, and V.v.nigrithorax;  
 RC TISSUE=Venom gland;  
 RA Shi W.U., Zhang S.F., Zhang C.X., Cheng J.A.;  
 RT "Cloning and sequencing of cDNA coding for prepro-melittin of *Polistes*  
 hebraeus, *Vespa magnifica* and *Vespa maculifrons*.";  
 Submitted (FEB-2002) to the EMBL/Genbank/DBS databases.  
 RL FUNCTION: Main toxin of bee venom with strong hemolytic activity.  
 CC -1- Integrates into cell membranes and has multiple effects, probably,  
 CC as a result of its interaction with negatively charged  
 CC phospholipids. It inhibits well known transport pumps such as the  
 CC Na(+)-K(+)-ATPase and the H(+)-K(+)-ATPase. Increases the  
 CC permeability of cell membranes to ions, particularly Na+ and  
 CC indirectly Ca(2+), because of the Na(+)-Ca(2+)-exchange (by  
 CC similarity).  
 CC -1- SUBUNIT: Monomer and homotetramer (by similarity).  
 CC -1- SUBCELLULAR LOCATION: Secreted (by similarity).  
 CC -1- TISSUE SPECIFICITY: Expressed by the venom gland.  
 CC -1- SIMILARITY: BELONGS TO THE MELLITIN FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; AF487907; AA012201.1; -;  
 DR EMBL; AF487910; AA012204.1; -;  
 DR EMBL; AF487908; AA012202.1; -;  
 DR Pfam; PF01372; Melittin; 1.  
 DR ProDom; PD014636; Melittin; 1.  
 KW Cytolysis; Hemolysis; Toxin; Signal; Amidation.  
 FT SIGNAL 1 21.  
 FT PROPEP 22 43 BY SIMILARITY.  
 FT CHAIN 44 69 REMOVED BY A DIPEPTIDYLPEPTIDASE (BY  
 FT MOD\_RES 69 MELITTIN).  
 FT FT AMIDATION (G-70 PROVIDE AMIDE GROUP) (BY  
 FT FT SIMILARITY).  
 SQ SEQUENCE 70 AA; 7543 MW; DA6B17C086C9560C CRC64;

Query Match 10.4%; Score 98.5; DB 1; Length 70;  
 Best Local Similarity 59.0%; Pred. No. 0.0048;  
 Matches 23; Conservative 4; Mismatches 7; Indels 5; Gaps 1;

QY 1 MKELVVALVFMVYISYIYAR-----LPDGITKAGEDA 34  
 DB 1 MKELVVALVFMVYISYIYARPEPPAPAEAEADAA 39

## RESULT 10

MEL\_POLHE

```

ID MEL POLHE STANDARD; PRT; 70 AA.
AC P59261;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Melittin precursor.
GN MELT.
OS Polistes hebraeus (Paper wasp).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Hymenoptera; Apocrita; Aculeata; Vespoidea;
OC Vespidae; Polistinae; Polistes.
OC NCBI_TaxId=202806;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Venom gland.
RA Shi W.J., Zhang S.F., Zhang C.X., Cheng J.A.;
RT "Cloning and sequencing of cDNA coding for prepro-melittin of Polistes
RT hebraeus, Vespina magnifica and Vespula maculifrons."
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: Has strong hemolytic activity. Integrates into cell
CC membranes and has multiple effects, probably as a result of its
CC interaction with negatively charged phospholipids. It inhibits
CC well known transport pumps such as the Na(+)-K(+)-ATPase and the
CC H(+)-K(+)-ATPase. Increases the permeability of cell membranes to
CC ions, particularly Na+ and indirectly Ca(2+), because of the
CC Na(+)-Ca(2+)-exchange (By similarity).
CC -1- SUBUNIT: Monomer and homotetramer (By similarity).
CC -1- SUBCELLULAR LOCATION: Secreted (By similarity).
CC -1- TISSUE SPECIFICITY: Expressed by the venom gland.
CC -1- SIMILARITY: BELONGS TO THE MELLITIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF487909; AAO12203.1; -
DR Pfam; PF01372; Melittin; 1.
DR ProDom; PD014636; Melittin; 1.
KM Cytolytic; Hemolysis; Toxin; Signal; Amidation.
FT SIGNAL 1 21
FT PROPEP 22 43
FT CHAIN 44 69
FT MOD_RRS 69 69
FT FT
FT MOD_RRS 69 69
SQ SEQUENCE 70 AA; 7513 MW; DA70167086C9560C CRC64;
Query Match 10.4%; Score 98.5; DB 1; Length 70;
Best Local Similarity 59.0%; Pred. No. 0.0048;
Matches 23; Conservative 4; Mismatches 7; Indels 5; Gaps 1;
QY 1 MKFLVNVAVLVFVYVYISYIYAR-----LPGGITKAGSDA 34
1 MKFLVNVAVLVFVYVYISYIYAR-----LPGGITKAGSDA 34
Db 1 MKFLVNVAVLVFVYVYISYIYAR-----LPGGITKAGSDA 39

```

```

OC Ephyrdoidea; Drosophilidae; Drosophila.
OX NCBI_TaxId=7227;
RN [1]
RP SEQUENCE FROM N.A. (ISOPFORM LONG).
RX MEDLINE=85267679; PubMed=2410772;
RA Mount S.M., Rubin G.W.;
RT "Complete nucleotide sequence of the Drosophila transposable element
RT copia: homology between copia and retroviral proteins."
RL Mol. Cell. Biol. 5:1630-1638(1985).
RN [2]
RP SEQUENCE FROM N.A., SEQUENCE OF 2-10, AND ALTERNATIVE SPLICING.
RX MEDLINE=85240569; PubMed=2409449;
RA Emori Y., Shiba T., Kanaya S., Inouye S., Yuki S., Saigo K.;
RT "The nucleotide sequences of copia and copia-related RNA in Drosophila
RT virus-like particles."
RL Nature 315:773-776(1985).
RN [3]
RP SEQUENCE FROM N.A. (ISOPFORM SHORT).
RX MEDLINE=89183629; PubMed=2538806;
RA Miller K., Rosenbaum J., Zbrzezna V., Pogo A.O.;
RT "The nucleotide sequence of Drosophila melanogaster copia-specific
RT 2.1-kb mRNA."
RL Nucleic Acids Res. 17:2134-2134(1989).
RN [4]
RP SEQUENCE FROM N.A. (ISOPFORM SHORT), AND MUTAGENESIS OF ASP-292.
RC TISSUE=Larva;
RX MEDLINE=90151630; PubMed=1689241;
RA Yoshioke K., Honma H., Zushi M., Kondo S., Togashi S., Miyake T.,
RA Shiba T.;
RT "Virus-like particle formation of Drosophila copia through
RT autocatalytic processing."
RL EMBO J. 9:535-541(1990).
CC -1- ALTERNATIVE PRODUCTS:
CC Name=Long;
CC IsoId=P04146-1; Sequence=Displayed;
CC Name=Short;
CC IsoId=P04146-2; Sequence=VSP 005226;
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY A11.
CC -1- SIMILARITY: Contains 1 CCHC-type zinc finger.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X04456; CAA28054.2; -
DR EMBL; X02599; CAA26444.1; -
DR EMBL; X02599; CAA26444.1; -
DR EMBL; X02600; CAA26446.1; -
DR EMBL; X02600; CAA26444.1; -
DR EMBL; X13719; CAA31997.1; -
DR EMBL; X54147; CAA38086.1; -
DR PIR; A03324; OFFPCP.
DR MEROPS; A11.001; -.
DR FlyBase; FBgn0013437; copia\GIP.
DR InterPro; IPR001969; Asprotease_site.
DR InterPro; IPR001584; Rve.
DR InterPro; IPR001878; Znf_CCHC.
DR Pfam; PF00665; rve; 1.
DR Pfam; PF00098; zf-CCHC; 1.
DR PRINTS; PR00939; C2HCZNFINGER.
DR SMART; SM00343; ZNF_C2HC; 1.
DR PROSITE; PS00141; ASP_PROTEASE; FALSE_NEG.
DR PROSITE; PS00158; ZF_CCHC; 1.
KM Transposable element_Hydrolyase; Aspartyl protease; ATP-binding;
KW Polypeptidein; Alternative splicing; Polymorphism; Zinc-finger.
FT CHAIN 1 270
FT CHAIN 271 1409
FT ZN_FING 230 247
CCHC-TYPE.

```

FT ACT SITE 292 292 PROTEASE (BY SIMILARITY).  
 FT VARSPPLIC 392 1374 Missing (in isoform short).  
 FT 1265 1288 /Frid=VSP\_005226.  
 FT VARIANT 1265 1288 STRGYLPRKEDPFLICMTKRRONS -> VOOGIYSKLLILI  
 FT 1289 1409 MISSING (IN VARIANT COPIA-RELATED).  
 FT MUTAGEN 1289 1409 D->A: LOSS OF ACTIVITY.  
 FT CONFLICT 191 191 S->N (IN REF. 2; CAA26447).  
 FT CONFLICT 300 300 I->V (IN REF. 2; CAA26447).  
 FT CONFLICT 866 866 O->R (IN REF. 2; CAA26447).  
 SQ SEQUENCE 1409 AA, 162817 MW, BE8944073A47691 CRC64;  
 Query Match 10.2%; Score 96.5; DB 1; Length 1409;  
 Best Local Similarity 23.8%; Pred. No. 0.25;  
 Matches 38; Conservative 27; Mismatches 48; Indels 47; Gaps 8;  
 QY 19 IYARLDGDTKAGEDLRPKS-----TAKHPFOED--NFCYIDNGKLPARGSVGN 70  
 DB 1016 IYARLDGDTKAGEDLRPKS-----TAKHPFOED--NFCYIDNGKLPARGSVGN 70  
 QY 71 MSRFVDPKADYGVGENTLV--HADDFEVPGESLKNM-----VRNLDVMPDIFET 119  
 DB 1066 VDRCIY--ILDKNINENIYLLVDDVYATGDMTRANNKRYLMEKFRMTDINEIKAF 1123  
 QY 120 LALRLVLDGVIMLRCPBLRVDYTSAYM-----WNQ 153  
 DB 1124 IGIRIEMQEDKIYL-----SOSAYVKIKLSKFNNE 1153  
 RESULT 12  
 ID VGLD HSV2 STANDARD; PRT; 393 AA.  
 AC P03172;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 21-JUL-1986 (Rel. 01, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Glycoprotein D precursor.  
 GN GD OR US6.  
 OS Herpes simplex virus (type 2).  
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
 OC Alphaherpesvirinae; Simplexvirus.  
 OX NCBI\_TaxID=10310;  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=84131549; PubMed=6321120;  
 RA Laaky L.A., Dowbenko D.J.;  
 RT "DNA sequence analysis of the type-common glycoprotein-D genes of  
 RL herpes simplex virus types 1 and 2.";  
 RL DNA 3:23-29(1984).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=333;  
 RX MEDLINE=84159516; PubMed=6323270;  
 RA Watson R.J.;  
 RT "DNA sequence of the Herpes simplex virus type 2 glycoprotein D  
 RL gene.";  
 RL Gene 26:307-312(1983).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BBKC;  
 RX Terhune S.S., Coleman K.T., Sekulovich R., Burke R.L., Spear P.G.;  
 RT Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.  
 RL -1- MISCELLANEOUS: THERE ARE SEVEN EXTERNAL GLYCOPROTEINS IN HSV1 AND  
 CC 2: GH, GB, GC, GD, GP, GL, AND GE.  
 CC -1- SIMILARITY: BELONGS TO THE HERPESVIRUSES GLYCOPROTEIN D FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; U12180; AAB60552.1; -  
 DR EMBL; U12182; AAB60554.1; -  
 DR EMBL; U12183; AAB60555.1; -  
 DR EMBL; K02373; AAA45842.1; -  
 DR EMBL; K01408; AAA45841.1; -  
 DR EMBL; AF021342; AAB72102.1; -  
 DR PIR; A03731; VGBED2.  
 DR InterPro; IPR002896; Herpes glycop.D.  
 DR Pfam; PF01537; Herpes glycop.D. 1.  
 KW Glycoprotein; Transmembrane; Signal.  
 FT SIGNAL 1 25  
 FT CHAIN 26 393 GLYCOPROTEIN D.  
 FT DOMAIN 26 339 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 340 363 POTENTIAL.  
 FT DOMAIN 364 393 CYTOPLASMIC (POTENTIAL).  
 FT CARBOHYD 119 119 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 146 146 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 287 287 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT VARIANT 249 249 T->I (IN STRAINS 333 AND BBKC).  
 FT VARIANT 337 337 A->S (IN STRAINS 333 AND BBKC).  
 FT VARIANT 367 369 REV->AQM (IN STRAINS 333 AND BBKC).  
 SQ SEQUENCE 393 AA, 43147 MW, A8514E21857AEDF2 CRC64;  
 Query Match 9.6%; Score 91; DB 1; Length 393;  
 Best Local Similarity 94.4%; Pred. No. 0.19;  
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 159 OPELAPEDPEDSALLIEDP 176  
 DB 290 OPELAPEDPEDSALLIEDP 307  
 RESULT 13  
 ID REPI\_ZYGRO STANDARD; PRT; 410 AA.  
 AC P13778;  
 DT 01-JAN-1990 (Rel. 13, Created)  
 DT 01-JAN-1990 (Rel. 13, Last sequence update)  
 DT 01-AUG-1990 (Rel. 15, Last annotation update)  
 DE Trans-acting factor B (REPI).  
 GN P.  
 OS Zygosaccharomyces rouxi (Candida mogii).  
 OG Plasmid PSRL.  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Zygosaccharomyces.  
 OX NCBI\_TaxID=4956;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=85210904; PubMed=3889347;  
 RA Araki H., Jeanpierre A., Tatsun H., Sakurai T., Ushio T.S.H.,  
 RA Muta T., Oshima Y.;  
 RT "Molecular and functional organization of yeast plasmid PSRL.";  
 RL J. Mol. Biol. 182:191-203(1985).  
 CC -1- FUNCTION: PLASMID PARTITION REQUIRE REPI, REP2, AND A CIS-ACTING  
 CC DNA SEQUENCE (KNOWN AS STB). REP 1 MAY ACT BY INTERCALATING IN  
 CC THE YEAST NUCLEAR MATRIX AND BINDING STB EITHER DIRECTLY OR VIA  
 CC REP2.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC EMBL; X02398; CAA26244.1; -  
 DR PIR; S28354; S28354.  
 KW Plasmid; Trans-acting factor.  
 SQ SEQUENCE 410 AA, 46578 MW, C9B28C381331F018 CRC64;  
 Query Match 8.6%; Score 81; DB 1; Length 410;







